Heterogeneity of Incipient Atrophy Patterns in Parkinson's Disease.

Pedro D. Maia^{3*}, Sneha Pandya^{1*}, Justin Torok¹, Ajay Gupta¹, Yashar Zeighami², Ashish Raj³

** equal contributions*

¹Department of Radiology, Weill Cornell Medical College of Cornell University
407 E. 61 Street, RRI06, New York, NY 10065, USA

² Montreal Neurological Institute, McGill University, Montreal, Canada

³Department of Radiology, Department of Radiology and Biomedical Imaging,
School of Medicine, University of California at San Francisco

Correspondence to: Ashish Raj, UCSF Radiology.
185 Berry Street Bldg B San Francisco, CA 94158, USA
E-mail: ashish.raj@ucsf.edu

Abstract

Parkinson's Disease (PD) is a the second most common neurodegenerative disorder after Alzheimer's disease and is characterized by cell death in the amygdala and in substructures of the basal ganglia such as the substantia nigra. Since neuronal loss in PD leads to measurable atrophy patterns in the brain, there is clinical value in understanding where exactly the pathology emerges in each patient and how incipient atrophy relates to the future spread of disease. A recent seed-inference algorithm combining an established network-diffusion model with an L1-penalized optimization routine led to new insights regarding the non-stereotypical origins of Alzheimer's pathologies across individual subjects. Here, we leverage the same technique to PD patients, demonstrating that the high variability in their atrophy patterns also translates into heterogeneous seed locations. Our individualized seeds are significantly more predictive of future atrophy than a single seed placed at the substantia nigra or the amygdala. We also found a clear distinction in seeding patterns between two PD subgroups – one characterized by predominant involvement of brainstem and ventral nuclei, and the other by more widespread frontal and striatal cortices. This might be indicative of two distinct etiological mechanisms operative in PD. Ultimately, our methods demonstrate that the early stages of the disease may exhibit incipient atrophy patterns that are more complex and variable than generally appreciated.

Keywords: Parkinson's Disease (PD); brain atrophy; neuropathology; tractography, seed-inference methods; network diffusion models.

Acknowledgements: The authors would like to thank C. Mezas and other members of the IDEAL lab for their insightful comments. AR and PM were supported by the NIH grant R01 EB022717. AR, SP, AG and JT were supported by the NIH grant R01NS092802.

Abbreviations: Parkinson's Disease (PD); Substantia Nigra pars Compacta (SN); Parkinson's Progressive Marker Initiative (PPMI); Network Diffusion Model (NDM);

1. Introduction

Parkinson's Disease (PD) is an age-related progressive neurodegenerative disorder clinically marked by tremor, bradykinesia, and rigidity. Patients exhibit cognitive deficits and with progression develop dementia, depression and/or other neuropsychiatric symptoms (Poewe and Wenning, 2006). Pathologically, these symptoms are caused by the loss of dopaminergic projections in the midbrain from the substantia nigra pars compacta (SN) to the striatum. A common histopathological hallmark of PD is the aggregation of misfolded α -synuclein (AS) protein called Lewy bodies (Jellinger, 2003). Histopathological studies of PD patient brains have strongly suggested that Lewy pathology spreads in a stereotypical stage-wise pattern from the brainstem to subcortical areas and then to the cerebral hemispheres along neuronal pathways (Braak *et al.*, 2003; Del Tredici and Braak, 2016). Recent advances in neuroimaging techniques have confirmed the Lewy pathology staging scheme. Regional MRI-derived atrophy was found to be the highest in SN, followed by atrophy of the striatum, amygdala, hippocampus and other limbic areas. Neocortical atrophy is seen in advanced stages, especially in orbitofrontal, insular and temporo-occipital cortices, in the same general order as seen in pathology stages (Ramírez-Ruiz *et al.*, 2005; Yau *et al.*, 2018).

Despite these advances, we lack understanding of pre-symptomatic stages of PD and it is unclear whether the pathological process begins in the brain or elsewhere in the nervous system. The purpose of this paper is to infer the likely sites of initiation of PD pathology from *in vivo* neuroimaging data, and to understand etiological origins and the sources of heterogeneity amongst PD populations. The basis of our approach is that the spread of PD pathology can be captured deterministically in a network-based model, which can

then be “inverted” to enable seed region inference. Here we briefly survey the neuropathological basis of this model and the seed-inference algorithm.

PD pathology progression occurs on the connectivity network

Toxic misfolded AS undergo template-driven aggregation in a “prion-like” fashion (Wood *et al.*, 1999; Yonetani *et al.*, 2009; Sacino *et al.*, 2013) intracellularly, followed by cell-to-cell transmission via trans-neuronal pathways along axonal projections to remote areas (Luk *et al.*, 2012; Masuda-Suzukake *et al.*, 2013; Rey *et al.*, 2016). A study showing a wild-type mice receiving a single intrastriatal injection of synthetic AS fibrils led to the trans-neuronal transmission of pathologic AS and Parkinson’s-like Lewy pathology in anatomically interconnected regions (Luk *et al.*, 2012). Lewy pathology spread occurs along the local and long-range fiber projections, thereby suggesting a process of “network spread”. This process is ubiquitous to neurodegenerative diseases, including Alzheimer’s, frontotemporal dementia, and others (Braak and Braak, 1991; Pandya *et al.*, 2017).

Recently, an *in vivo* quantitative approach using predictive network diffusion model (NDM) has been applied to AD and other dementias by specifically modeling the trans-neuronal spread using connectivity (Raj *et al.*, 2012, 2015). The diagnostic and clinical role of NDM and other models of neurodegenerative pathologies through trans-neuronal diffusion is detailed in (Iturria-Medina, 2013; Iturria-Medina *et al.*, 2014; Carbonell *et al.*, 2018). Likewise, a network-based analysis has been applied to a large group of PD patients showing atrophy distribution by modeling the trans-neuronal spread of AS outward from SN (Zeighami *et al.*, 2015). The study was statistically validated using neuropathological and neuroimaging data of 232 PD patients which demonstrated the spatial and temporal atrophy patterns in PD reflecting out of the disease epicenter in the SN. The NDM was validated on the same data series in our laboratory, where we showed that SN-seeded NDM was able to faithfully recapitulate PD cross-sectional atrophy patterns, both at the group level and at the individual level; the latter at lower predictive power and higher variability (Pandya *et al.*, 2018), providing a new computationally-defined staging scheme that mirrors Braak’s six-stage Lewy pathology staging scheme (Braak *et al.*, 2003).

Utilizing the Network Diffusion model for seed inference

In this paper we report that the NDM can be successfully “inverted” in order to infer the most likely pattern of incipient pathology seeding from available regional atrophy data in PD patients. For this purpose, we

leverage a recently developed seed-inference algorithm for Alzheimer's pathology and tailor it to the parkinsonian context. Torok et al., 2018 combined a network diffusion model that successfully recapitulates patterns of regional brain atrophy (Raj et al., 2012) with an L1-penalized optimization routine to infer the origins of pathology across individual subjects from the Alzheimer's Disease Neuroimaging Initiative (ADNI) public database. The key requirement of this seed-inference algorithm is a *forward* model, which can predict, from a given patient's baseline seed pattern their ongoing progression of longitudinal atrophy. Since we have already shown that the spatiotemporal dynamics of PD pathology spread can be given quantitatively and deterministically from the NDM (Pandya et al., 2018), we chose it as the forward model. Following Torok et al., 2018, we then design an inference algorithm which minimizes a cost function involving the forward model, and an additional penalty term that penalizes non-sparse solutions. Hence, our algorithm optimizes two constraints: (i) the *seed pattern, when extrapolated to future time*, must match the patient's observed regional atrophy values, and (ii) the *seed pattern must be sparse*, such that only a small set of incipient regions will exhibit early vulnerability.

We apply our seed inference algorithm to a large public study of PD patients called the Parkinson's Progressive Marker Initiative (PPMI) (www.ppmi-info.org/data). The same cohort was used in our previous NDM study (Pandya et al., 2018), with 232 patients diagnosed with PD along with 117 age-matched controls (see **Table 1**). On this data we demonstrate that the group average pattern of inferred seeding is in strong agreement with known early sites in the brainstem, striatum and limbic areas. Intriguingly, however, there is considerable individual variability in seeding, suggesting that a common incipient state of neurodegeneration cannot explain the intersubject variability observed empirically. Thus, our methods demonstrate that the early stages of the disease may exhibit seeding patterns that are more complex and variable than generally appreciated. We identified several potential prodromal seeding variants based on clustering analysis, which, if verified in future studies, could reveal new insights into the sources of etiologic variability of PD.

Table 1.

Caption: Demographic and clinical details of PPMI cohort. MoCA = Montreal Cognitive Assessment, UPDRS = Uniform Parkinson's disease Rating Score

	Male	Female	Age	MoCA	UPDRS3	Hoch and Yarh stage
HC (n= 117)	74	43	59.5 ± 11.3	28.2 ± 1.2	-	-
PD (n = 232)	155	77	61.2 ± 9.1	27.3 ± 2.2	21.9 ± 9.1	1.6 ± 0.5

2. Materials and Methods

2.1 Participants and longitudinal MRI morphometric processing

A cross-sectional imaging dataset from PPMI was used to test our model for current study. 3T high-resolution T1-weighted MRI images from the baseline visit of 232 de-novo Parkinson's patients as well as 117 age matched Healthy controls (HC) were obtained from PPMI database to measure PD related regional atrophy. Demographic and clinical features of our study are show in **Table 1** at the time of the study. The images were denoised, normalized, and corrected for non-uniformity intensity. All images were first linearly and then non-linearly registered to the MNI-ICBM 152 template. Regional brain atrophy was computed using deformation-based morphometry (DBM) (i.e. the determinant of the Jacobian transformation matrix obtained from nonlinear transformation fields). We extracted mean regional DBM values per subjects after calculating voxel-wise DBM maps. A two tailed t-test was performed between PD and HC mean DBM values to calculate the effect size of PD related regional atrophy which is represented by a vector (N=112). This t-statistic was converted to the natural range [0,1] using the logistic transform, following (Raj *et al.*, 2015). These atrophy measures were then used to test the modeling analyses.

For this study, we used a brain parcellation with 112 regions whose details have been reported previously (Zeighami *et al.*, 2015). Thirty-four cerebellar regions were removed, leaving 78 cerebral regions. In brief, supratentorial regions i.e. cortical and basal ganglia related regions were included from Hammers atlas (Hammers *et al.*, 2003). Furthermore, three midbrain structures, subthalamic nucleus, substantia nigra, and red nucleus were manually segmented using the high-resolution MRI template (T1-weighted ICBM152 template, resolution = 0.5 mm³), the BigBrain (Amunts *et al.*, 2013), and the brainstem anatomical atlas of (Duvernoy, 1995). A subcortical atlas based on ultrahigh-field MRI (Keuken *et al.*, 2014) was used to confirm the accuracy of the segmentations.

2.2 White matter connectome

An Illinois Institute of Technology Human Brain atlas v.3 constructed from high-resolution diffusion weighted MRI (DW-MRI) data from 72 young healthy subjects was used for structural connectivity. We used the anatomical connection density (ACD) as the measure of connectivity for this paper which is defined as the fraction of the connected superficial nodes with respect to the total number of superficial nodes of both areas, as proposed in (Iturria-Medina *et al.*, 2007). ACD accounts for correcting varying brain region sizes as it is obtained by dividing the raw connection strength value by the sum of region-pair surface areas. We refer to this network using the *connectivity matrix* $C = \{c_{i,j}\}$ whose elements $c_{i,j}$ represent the connection strength of white matter fiber pathways between i^{th} and j^{th} gray matter regions. Connections are assumed to be bidirectional due to limitations of the DTI tractography data.

2.3 Predictive PD Network Diffusion Model

Raj *et al.* (2012) demonstrated that the spread of proteinopathic agents over time is well captured by a dynamical system defined over a network-graph rendering of the brain, with the nodes representing gray matter structures and inter-regional connections defined as above. The *forward* Network Diffusion Model (NDM) for the pathology load $\mathbf{x}$ at time t is given by

$$\frac{d\mathbf{x}}{dt} = -\beta L \mathbf{x}(t),$$

where L is the Graph Laplacian matrix given by

$$L_{i,j} = \begin{cases} -c_{i,j}, & \text{for } i \neq j \text{ and } c_{i,j} \neq 0, \\ \sum_{t,j \in \varepsilon} c_{t,j}, & \text{for } i = j, \\ 0, & \text{otherwise.} \end{cases}$$

The NDM has a closed form analytical solution,

$$\mathbf{x}(t) = e^{-\beta L t} \mathbf{x}_0,$$

where $\mathbf{x}_0 = \mathbf{x}(t = 0)$ is referred to as the seed vector or incipient atrophy pattern on which the diffusion kernel $e^{-\beta L t}$ acts as a temporal and spatial blurring operator for the connectivity matrix C . In general, the unit of model's diffusion time t is arbitrary and global diffusivity β is unknown. We chose a value that would roughly span Parkinson's progression (10-20 years), giving $\beta = 0.15$. Thus, for a given initial atrophy configuration, we can use the NDM to predict the atrophy pattern at all future time points. We

have previously validated the accuracy of the NDM predictions in multiple contexts (Raj *et al.*, 2012, 2015; Pandya *et al.*, 2017; Raj and Powell, 2018).

Glossary of symbols

Symbol	Description
$\mathbf{b}$	Empirically-observed atrophy vector
$\langle \mathbf{b} \rangle_{PD}$	Average of $\mathbf{b}$ across all PD patients
$\mathbf{x}_{seed}$	Seed vector, incipient atrophy pattern
$\langle \mathbf{x}_{seed} \rangle_{PD}$	Average of $\mathbf{x}_{seed}$ across all PD patients
$\mathbf{a}(t, \mathbf{x}_{seed})$	Predicted atrophy vector via NDM

2.4 Seed-inference method

Here we recapitulate some of the key points of the seed inference method from Torok *et al.*, 2018. The empirical atrophy z-scores are given by

$$z_{ij} = (v_{ij} - \mu_{i,control}) / \sigma_{i,control}$$

where the indexes denote the i -th brain region and j -th patient respectively; μ and σ denote mean and standard deviation calculated with respect to the healthy controls. The scores were then normalized by a weighted logistic transform to keep values within the (0,1) range; these normalized vectors will be referred to in what follows as *empirically-observed atrophy vectors* $\mathbf{b}$ against which we run our inference algorithm (Raj *et al.*, 2012; Torok *et al.*, 2018).

The forward NDM can be used to infer the most likely pattern of disease seeding $\mathbf{x}_{seed}$ from a given vector $\mathbf{b}$. This inverse seed-inference process utilizes a constrained optimization algorithm with a L_1 -penalized cost-function to promote sparsity while maximizing the Pearson correlation R between the NDM-predicted atrophy vector $\mathbf{a}(t, \mathbf{x}_{seed})$ and $\mathbf{b}$.

Analogous to the original formulation of the NDM, we define the two-variable function $\mathbf{a}$ as:

$$\mathbf{a}(t, \mathbf{x}_{seed}) = e^{-\beta L t} \mathbf{x}_{seed}$$

Following Torok *et al.*, 2018, we perform the minimization over the two active variables, t and $\mathbf{x}_{seed}$, in two steps. After determining an initial guess for the seed pattern, $\mathbf{x}_{seed}^*$, and running the NDM on each region individually, we find an estimate for the t_{min} that satisfies the following condition:

$$t_{min} = \underset{t}{\operatorname{argmin}}(e^{-R(\mathbf{a}(t, \mathbf{x}_{seed}^*), \mathbf{b})})$$

where $R(\mathbf{a}(t, \mathbf{x}_{seed}^*), \mathbf{b})$ is the Pearson correlation between the NDM-predicted atrophy using initial atrophy $\mathbf{x}_{seed}^*$ and the *empirically-observed atrophy vector* $\mathbf{b}$. This cost function is monotonically decreasing and is minimized when the Pearson correlation is maximized, which is consistent with the NDM criterion. We then use this estimate of t_{min} to find the $\mathbf{x}_{seed}$ that minimizes the following L_1 -constrained cost function:

$$\operatorname{cost}(\mathbf{x}_{seed}) = e^{-R(\mathbf{a}(t_{min}, \mathbf{x}_{seed}), \mathbf{b})} + \lambda \|\mathbf{x}_{seed}\|_1$$

λ is a tunable parameter that controls the sparsity of solutions; higher values of λ force more entries of the resultant $\mathbf{x}_{seed}$ vector to 0. We used the same stopping criteria as Torok *et al.*, 2018, and convergence under these conditions was not an issue in all runs we observed. See the original publication for further algorithmic details.

3. Results

For every subject in our study, we infer an individualized seed vector $\mathbf{x}_{seed} \in \mathbb{R}^{78}$ that minimizes the cost function above. Each entry of the seed vector is associated with a specific gray matter region given by the brain atlas, and represents the likelihood that proteinopathic agents are present at that region when $t = 0$. When projected forward in time by the NDM, the sparse seed vectors yield predicted atrophy vectors $\mathbf{a}(t, \mathbf{x}_{seed})$ that are highly correlated with their corresponding empirically-observed atrophy vectors $\mathbf{b}$. The entries of vectors $\mathbf{a}$ and $\mathbf{b}$ will be referred to as predicted and observed atrophy-values respectively. Finally, we will use brackets and subscripts of the form $\langle \text{argument} \rangle_{\text{subgroup}}$ to indicate the average of vectors across subjects from a given subgroup (i.e. HC and PD).

3.1 Group-level regional patterns for atrophy and seed vectors

Figures 1A-C show the axial, coronal, and sagittal glassbrain views respectively of $\langle \mathbf{b} \rangle_{PD}$, i.e., the average of the empirically-observed atrophy vectors across all PD patients. **Figures 1D-F** show the average of the inferred seed vectors $\langle \mathbf{x}_{seed} \rangle_{PD}$ across all PD patients. We note that $\langle \mathbf{x}_{seed} \rangle_{PD}$ differs from the seed vector inferred directly from $\langle \mathbf{b} \rangle_{PD}$, since the averaging and seed-inferencing mathematical operations do not commute. In all panels, the load of proteinopathic agents in a brain region is proportional to the sphere diameter placed there, although actual values were scaled for improving visualization. **Table 2** compares the top-10 brain regions associated with the largest entries of $\langle \mathbf{b} \rangle_{PD}$ with those of $\langle \mathbf{x}_{seed} \rangle_{PD}$. There is significant overlap between the top-10 regions for both vectors, especially regarding the Putamen, Pallidum, and Red Nucleus.

Fig. 1: Visualization of average atrophy vector and inferred seed vector for PD patients. Panels show (A) the axial, (B) the coronal, and (C) the sagittal glassbrain views respectively of the average of the empirically-observed atrophy vectors across all PD patients $\langle \mathbf{b} \rangle_{PD}$. Panels (D-F) show the same views for the average seed vectors across all PD patients $\langle \mathbf{x}_{seed} \rangle_{PD}$. The pathology load in a brain region is proportional the sphere diameter placed on it. Actual values of atrophy and seed were scaled for improved visualization

Table 2.

Caption: (Left) Brain regions associated with the largest entries of the empirically-observed atrophy vector averaged across all PD patients $\langle \mathbf{b} \rangle_{PD}$. (Right) Brain regions associated with the largest entries of the inferred seed vector averaged across all PD patients $\langle \mathbf{x}_{seed} \rangle_{PD}$

Rank	Brain regions associated with largest entries of $\langle \mathbf{b} \rangle_{PD}$	Value	Brain regions associated with largest entries of $\langle \mathbf{x}_{seed} \rangle_{PD}$	Value
1	Putamen (right)	.15	Red Nucleus (left)	.30
2	Pallidum (left)	.15	Putamen (right)	.25
3	Pallidum (right)	.15	Pallidum right	.22
24	Precentral gyrus (left)	.15	Substantia Nigra (right)	.21
5	Red Nucleus (left)	.15	Red Nucleus (right)	.20
6	Nucleus accumbens (left)	.14	Subthalamic Nucleus (left)	.20
7	Putamen (left)	.14	Substantia Nigra (left)	.19
8	Thalamus (right)	.14	Pallidum (left)	.18
9	Straight gyrus (right)	.14	Pre-subgenual frontal cortex (left)	.18
10	Inferior frontal gyrus (left)	.14	Precentral gyrus (left)	.17

3.2 L-curves and parameter choices

Torok *et al.*, 2018 presented a novel method to determine a consistent seeding pattern for each patient. Naturally, the inverse-problem algorithm requires a sensible choice of the L_1 - penalty parameter λ . A priori, one does not know what this sensible choice should be, thus, it is customary to explore the parameter space within a reasonable range of values (0.1: 0.05: 1) and analyze the corresponding L-curves. The goal is to select the "elbow" of the curve, i.e., a value that provides a sensible tradeoff between the mismatch of model/data – given by the $e^{-R(\mathbf{a}(t_{min}, \mathbf{x}_{seed}), \mathbf{b})}$ term in the cost-function, and the sparsity of the seed, given by the $\lambda \|\mathbf{x}_{seed}\|_1$

Figure 2 shows how the mismatch between model/data decreases as we (i) increase the value of λ , (ii) increase the number of seeds, or (iii) increase the L_1 norm of the seed. We argue that $\lambda = 0.25$ can be considered the "elbow" for the black curves; it provides an intermediate value of 0.63 (it is higher than the min = 0.52 and lower than the max 0.74), an intermediate average number of seeds of 6 (higher than 1 and lower than 19), and an intermediate L_1 - norm of 1 (higher than .2 and lower than 1.5). The curves in black are average values (± 1 standard deviation) of the individualized inferred seeds. The curves in red correspond to the inferred seed from the population average atrophy-pattern. The discrepancy between curves for higher lambda values indicate the steps of the algorithm do not commute, i.e., that finding seeds and calculating their average (black) will lead to results that are significantly different than averaging the atrophies and inferring a seed from that averaged value. This strongly advocates for seed inference using patient-specific atrophy patterns as opposed to population averages, although the two curves are not too far apart for our selected value of $\lambda = 0.25$. All inferred seeds presented in our plots and tables were found using this value.

Fig. 2: L-curves for parameter calibration. The rationale behind our choice for lambda is to make a sensible tradeoff between low mismatch between model/data and a high-sparsity for the inferred seed vector. The plots above show how the mismatch (exponential term defined in the cost function) varies in terms of (A) the lambda value, (B) the number of non-zero entries in the seed vector, and (C) the L1-norm of the seed vector. The curves in black are average values (± 1 standard deviation) of the individualized inferred seeds. The curves in red correspond to the inferred seed from the average atrophy pattern. The discrepancy between curves for higher lambda values indicate the steps of the algorithm do not commute, i.e., that finding seeds and calculating their average (black) will lead to results that are significantly different than averaging the atrophies and inferring a seed from that averaged value. Notice that an increase in sparsity is correlated with an increase in the data/model mismatch and vice-versa. Thus, we

choose $\lambda = 0.25$ as it yields both a tractable number of non-zero entries in the seeds vector (6 out of 78) and a tolerable mismatch (value = 0.63)

3.3 Individualized seeds are more predictive of future atrophy

The inferred $\mathbf{x}_{\text{seed}}$ vectors varied significantly from patient to patient. In this subsection, we demonstrate that inferred $\mathbf{x}_{\text{seed}}$ are more predictive of atrophy pattern than a common seed located at a single region, i.e., a seed pattern defined by a vector with only one non-zero entry placed either in the Amygdala or in the Substantia Nigra. While certain brainstem nuclei, might be better candidate seed locations of PD pathology (e.g. medulla oblongata), these regions are not accessible on MRI and therefore we have focused on seed locations corresponding to Braak stage-III and higher.

The algorithm is as follows:

1. Infer an individualized $\mathbf{x}_{\text{seed}}$ from the observed atrophy vector $\mathbf{b}$ using the *inverse model*.
2. Project each $\mathbf{x}_{\text{seed}}$ into the future using the forward model to obtain the predicted atrophy vector $\mathbf{a}(t, \mathbf{x}_{\text{seed}})$. Project the single-seed vector $\mathbf{x}_{\text{single}}$ the same way to obtain $\mathbf{a}(t, \mathbf{x}_{\text{single}})$.
3. Obtain the Pearson's correlation coefficient (R-max) between (i) the pair of vectors $\{\mathbf{a}(t, \mathbf{x}_{\text{seed}}), \mathbf{b}\}$, (ii) the pair of vectors $\{\mathbf{a}(t, \mathbf{x}_{\text{single, Subst. Nigra}}), \mathbf{b}\}$, and (iii) the pair of vectors $\{\mathbf{a}(t, \mathbf{x}_{\text{single, Amygdala}}), \mathbf{b}\}$, for each PD patient. Create histograms for R-max for all cases. See **Figure 3**.

Figures 3A-C show the R-max histograms (i)-(iii) defined above in black, green, and in magenta respectively. The R-max values associated with (i) are typically higher than the ones associated with (ii) and (iii), demonstrating that inferred individualized $\mathbf{x}_{\text{seed}}$ patterns lead to significantly more predictive patterns of the patients' atrophy vectors than a common single-seed vector. This result provides strong evidence against a stereotyped/standard single seed location. We also find that (**Fig. 3C**) that a single seed in the Substantia Nigra is more likely than a single seed in the Amygdala. Of course, if the seed inference algorithm was giving trivial outcomes (e.g. inferred seed pattern = observed atrophy pattern) then we would erroneously obtain similar results to the above. To guard against that possibility, In **Fig. 3D**, we present a R-max histogram comparing the two vectors $\{\mathbf{b}, \mathbf{x}_{\text{seed}}\}$, showing that $\mathbf{x}_{\text{seed}}$ are not obvious

correlates of the observed atrophy patterns. This is consistent with the complex dynamics of disease spread in PD and suggests that our seed inference is implicating a different set of regions than would be trivially predictable from the most atrophied regions.

Fig. 3: Individualized seeds are significantly more predictive of future atrophy. Our forward model estimates the progression of atrophy patterns of a given patient starting from the inferred seed pattern. Our individualized seeds lead to future atrophy patterns that are more similar to the observed ones (higher R_{max} values, in black histograms) than the ones led by a single seed placed at (A) the substantia nigra, or (B) the amygdala. This result provides strong evidence against a stereotyped/standard single seed location, although (C) shows that a single seed in the substantia nigra is more likely than one in the amygdala. (D) shows that our inferred seed patterns are not obvious correlates of the observed atrophy patterns, which is consistent with the complex dynamics of disease spread in PD.

3.4 Hierarchical clustering and PD subgroups

In this section we demonstrate that despite the heterogeneity of the incipient atrophy patterns across subjects, the inferred seed vectors $\mathbf{x}_{seed}$ can be categorized in two subgroups (S1 and S2). Analogously, the empirically-observed atrophy vectors $\mathbf{b}$ can also be categorized in two subgroups (A1 and A2). We use a standard technique in data mining known as standard agglomerative hierarchical cluster tree to determine, data-driven clusters in the data based on subject's dissimilarity. In our analysis, we used MATLAB's *linkage* function (to compute distance) with inner squared distance (Ward) and Euclidean metric.

Figures 4A-B summarize the sizes of all four subgroups: S1 (65 patients in cyan) vs S2 (167 patients in black) for seeds & A1 (99 patients in red) vs A2 (133 patients in blue) for atrophies. Notice that the majority of patients classified in S2 (100/167) are also classified in A2, but there is significant mixing between seed and atrophy subgroups. Visual inspection of the histograms in **Figures 4C-E** reveal that patients in all subgroups exhibit significant overlap regarding their age, and MoCA/ UPDRS3 scores.

A

B

C

Fig. 4: PD subgroups for seeds and atrophy. (A) We use MATLAB's agglomerative hierarchical cluster tree function (linkage) to categorize the seed and atrophy data into two subgroups each: S1 (65 patients in cyan) vs S2 (167 patients in black) for seeds & A1 (99 patients in red) vs A2 (133 patients in blue) for atrophies. Distances between clusters were calculated using the inner square distance (ward) and the euclidean metric. (B) Classification of the 232 PD patients in our study following the color scheme above (A1 vs A2 on top, S1 vs S2 on bottom). Histograms normalized by probability for (C) age, (D) MoCA scores, and (E) UPDRS3 scores for each subgroup. The histograms exhibit significant overlap for all variables. Using single linkage distance (nearest neighbor) instead of the ward distance does not lead to subgroups of significant sizes as it only discriminates a few outliers. See Tables 4-5 for a detailed description of the atrophy and seed patterns for each subgroup.

367 **Figure 5** shows on the top of each panel, a glassbrain view of the inferred seed vectors $\langle \mathbf{x}_{seed} \rangle_{S1}$ and
 368 $\langle \mathbf{x}_{seed} \rangle_{S2}$ averaged between subjects in the S1 (cyan) and in the S2 (black) subgroups respectively. In
 369 the middle of each panel, we show a glassbrain view of AFS1 (cyan) and AFS2 (black). They represent
 370 the average predicted atrophy pattern (via NDM) of the seeds classified in S1 and S2 subgroups, i.e., $\langle$
 371 $\mathbf{a}(t, x_{seed} \in S1) \rangle$ and $\langle \mathbf{a}(t, x_{seed} \in S2) \rangle$ respectively. Finally, in the bottom of each panel, we
 372 show a glassbrain view of $\langle \mathbf{b} \rangle_{A1}$ and $\langle \mathbf{b} \rangle_{A2}$, the empirically-observed atrophy vectors averaged
 373 between subjects in A1 (red) and A2 (blue) subgroups respectively.

374

375 **Fig. 5:** Glassbrains of the average seed pattern, projected atrophy, and observed atrophy in two
 376 subgroups found using hierarchical clustering. (A) Top to bottom: S1 and S2 are the individualized seed
 377 vectors averaged across subgroups S1 and S2, AFS1 and AFS2 are average projected atrophy for

subgroups $S1$ and $S2$ derived from forward NDM, and $A1$ and $A2$ are observed atrophy from subgroups $A1$ and $A2$. Distinction of two subgroups from seed pattern as seen in $S1$ and $S2$ is evident as opposed to projected atrophy ($AFS1$, $AFS2$) and observed atrophy ($A1$, $A2$). We can clearly identify two different patterns, $S1$ showing dominant seed distribution in midbrain regions and $S2$ showing widespread cortical distribution. (B) and (C) shows glass brain visualization in coronal and sagittal view respectively.

Table 3 compares the top-10 brain regions associated with the largest entries of $\langle \mathbf{x}_{seed} \rangle_{S1}$ with those of $\langle \mathbf{x}_{seed} \rangle_{S2}$, and **Table 4** shows the top-10 brain regions associated with those of $\langle \mathbf{b} \rangle_{A1}$ and $\langle \mathbf{b} \rangle_{A2}$. Finally, **Table 5** shows the regions associated with the largest entries of $\langle \mathbf{a}(t, \mathbf{x}_{seed} \in S1) \rangle$ and $\langle \mathbf{a}(t, \mathbf{x}_{seed} \in S2) \rangle$ respectively. These results are based on the hierarchical clustering analysis explained previously at two cluster level for both seed and atrophy data.

Table 3.

Caption: The individual patient-dependent seed vectors $\mathbf{x}_{seed}$ can be categorized in two subgroups ($S1$ and $S2$) via agglomerative hierarchical cluster tree method. Left column shows the brain regions associated with the largest entries of $\langle \mathbf{x}_{seed} \rangle_{S1}$ (in cyan) and right column shows the brain regions associated with the largest entries of $\langle \mathbf{x}_{seed} \rangle_{S2}$ (in black).

Rank	Brain regions associated with largest entries of $\langle \mathbf{x}_{seed} \rangle_{S1}$	Value	Brain regions associated with largest entries of $\langle \mathbf{x}_{seed} \rangle_{S2}$	Value
1	Red Nucleus (left)	.49	Putamen (right)	.34
2	Red Nucleus (right)	.34	Precentral gyrus (left)	.25
3	Subthalamic Nucleus (left)	.33	Lingual gyrus (right)	.22
4	Substantia Nigra (left)	.31	Pre-subgenual Frontal cortex (left)	.22
5	Substantia Nigra (right)	.30	Lingual gyrus (left)	.21
6	Pallidum (left)	.26	Inferior frontal gyrus (left)	.20
7	Subthalamic Nucleus (right)	.26	Cuneus (right)	.19
8	Thalamus (right)	.20	Superior parietal Gyrus (left)	.18
9	Pallidum (right)	.20	Cuneus (left)	.18
10	Thalamus (left)	.15	Subgenual frontal cortex (left)	.18

Table 4.

Caption: The individual, empirically-observed atrophy vectors $\mathbf{b}$ can be categorized in two subgroups (A1 and A2) via agglomerative hierarchical cluster tree method. Left column shows the brain regions associated with the largest entries of $\langle \mathbf{b} \rangle_{A1}$ (in red) and right column shows the brain regions associated with the largest entries of $\langle \mathbf{b} \rangle_{A2}$ (in blue).

Rank	Brain regions associated with largest entries of $\langle \mathbf{b} \rangle_{A1}$	Value	Brain regions associated with largest entries of $\langle \mathbf{b} \rangle_{A2}$	Value
1	Nucleus accumbens (left)	.16	Precentral gyrus (left)	.16
2	Putamen (left)	.16	Inferior frontal gyrus (right)	.16
3	Putamen (right)	.16	Inferior frontal gyrus (left)	.15
4	Thalamus (right)	.16	Posterior temporal lobe (right)	.14
5	Pallidum (left)	.16	Pallidum (left)	.14
6	Pallidum (right)	.16	Pallidum (right)	.14
7	Medial orbital gyrus (right)	.15	Straight gyrus (left)	.14
8	Red Nucleus (left)	.15	Red Nucleus (left)	.14
9	Substantia Nigra (right)	.15	Substantia Nigra (left)	.14
10	Subthalamic Nucleus (left)	.15	Substantia Nigra (right)	.14

Table 5.

Caption: The individual inferred-seed vectors $\mathbf{x}_{seed}$ can be projected forward in time (via NDM) to predict the future atrophy patterns $\mathbf{a}(t, \mathbf{x}_{seed})$. The AFS1 (cyan) and AFS2 (black) represent the average predicted atrophy pattern of the seeds classified in S1 and S2 subgroups, i.e., $\langle \mathbf{a}(t, \mathbf{x}_{seed} \in S1) \rangle$ and $\langle \mathbf{a}(t, \mathbf{x}_{seed} \in S2) \rangle$ respectively. Left/Right column show the brain regions associated with the largest entries of each vector.

Rank	Brain regions associated with largest entries of	Value	Brain regions associated with largest entries of	Value
------	--	-------	--	-------

	$\langle a(t, x_{seed} \in S1) \rangle$		$\langle a(t, x_{seed} \in S2) \rangle$	
1	Thalamus (right)	.19	Precentral gyrus (left)	.16
2	Red Nucleus (left)	.19	Inferior frontal gyrus (right)	.15
3	Substantia Nigra (left)	.19	'Presubgen_antCing_I'	.15
4	Substantia Nigra (right)	.19	Nucleus accumbens (left)	.14
5	Subthalamic Nucleus (left)	.19	Putamen (right)	.14
6	Anterior temporal, lobe, medial part (left)	.18	Straight gyrus (right)	.14
7	Pallidum (right)	.18	Inferior frontal gyrus (left)	.14
8	Red Nucleus (right)	.18	Medial orbital gyrus (right)	.14
9	Putamen (left)	.17	Lateral remainder of occipital lobe (left)	.13
10	Putamen (right)	.17	Middle frontal gyrus (right)	.13

408

409 4. Discussion

410 The Braak staging model of PD, based on large post-mortem autopsy series, proposes that the deposition
411 of misfolded AS occurs in a stereotypical fashion: early sites of pathological involvement are the
412 brainstem (stages 1-2), SN and limbic structures (stages 3-4), followed by wider spread into adjoining
413 frontal and other cortices (stages 5-6). Loss of dopaminergic neurons in the SN and other cerebral
414 structures constitutes in an extensive neurodegenerative process (Del Tredici *et al.*, 2002; Braak *et al.*,
415 2003). Despite these stereotypical patterns, there is strong evidence of inter-subject variability in etiology
416 and progression of PD patients. It is becoming increasingly evident that many patients cannot be fitted
417 into the canonical progression scheme (Burke *et al.*, 2008; Rietdijk *et al.*, 2017). Since antemortem
418 imaging cannot reveal directly the incipient patterns of pathology, this study aims to develop a principled,
419 model-based method of “looking back” into the disease trajectory based on present imaging data. Many
420 clinical uses of this approach are enumerated below, but the key outcome is to enable assessment of a

patient's incipient seeding pattern, and to use this pattern to cluster them into clinically relevant subtypes. Olfactory impairment, dementia, depression and other neuropsychiatric symptoms appear in preclinical stages of PD, preceding motor manifestations by long periods (Shiba *et al.*, 2000; Schuurman *et al.*, 2002; Leentjens *et al.*, 2003; Ross *et al.*, 2006, 2008). It is therefore plausible that the ability to infer seeding patterns of a patient can reveal likely sources of symptomatic heterogeneity.

In this paper we leverage a recently developed seed-inference algorithm for Alzheimer's pathology (Torok *et al.*, 2018) and tailor it to the parkinsonian context. Torok *et al.*, 2018 combined a network-diffusion model that successfully recapitulates patterns of regional brain atrophy (Raj *et al.*, 2012) with an L1-penalized optimization routine to infer the likely origins of pathology across individual subjects from the Alzheimer's Disease Neuroimaging Initiative (ADNI) public database. Their results showed that the high degree of variability between patients at baseline translates to even more heterogeneous seed patterns that are significantly more predictive of future atrophy than a single seed placed in the hippocampus. Given that successful graph-theoretic models are also available for Parkinson's Disease (Zhou *et al.*, 2012; Zeighami *et al.*, 2015; Yau *et al.*, 2018), it is sensible to investigate if similar methods are capable of revealing seeding heterogeneity in PD. In this study we found that the observed heterogeneity in PD atrophy is best explained by the heterogeneous seeding patterns. Furthermore, these seeding patterns are clustered into two main sub-groups which are not obvious from their observed atrophy. We also confirmed that inferred seeding patterns are more predictive of future atrophy patterns than a single seed placed at commonly-found sites of neuronal loss, further cementing the role of etiologic heterogeneity in PD patients. Below we discuss each of these key results in context and describe their implications.

Group atrophy patterns of PPMI cohort

Figure 1 (top) shows glassbrain views for the empirically-observed atrophy vectors averaged across all PD patients $\langle b \rangle_{PD}$. As shown in **Figures 1** and in **Table 2**, the region with the highest atrophy value is the Putamen; its primary role is to regulate movement at various stages. It plays an important role in PD because its inputs and outputs are interconnected to the Substantia nigra and Pallidus (DeLong and Wichmann, 2007). The pallidum is the next region with more atrophy.

Group average inferred seeding patterns implicate common sites of early synuclein pathology and Braak staging

Figure 1 (bottom) shows the average of the inferred seed vectors $\langle \mathbf{x}_{\text{seed}} \rangle_{PD}$ from their atrophy profiles. The main regions implicated as the most likely early seeding sites in **Table 1** (SN, RN, and striatal areas like putamen and accumbens) are all regions where early PD pathology is observed (Braak *et al.*, 2003; Huot and Parent, 2007; Hanganu *et al.*, 2014; Lewis *et al.*, 2016) They also highlight a prominent role for the red nucleus in the rostral midbrain as potential *source* of the disease. It exhibits a higher seed value than regions commonly associated to PD such as the Putamen, Pallidum and Substantia Nigra. This is consistent with the notion that PD does not necessarily being in the SN (Del Tredici *et al.*, 2002; Braak *et al.*, 2003; Lang and Obeso, 2004; Ahlskog, 2005), but could also be due to the difficulty of disambiguating fiber projections to and from SN versus those that terminate at RN. Brain regions such as the Thalamus and the Nucleus accumbens were not among the top-10 seed locations despite their high atrophy values. Instead, larger seed values were found in the right Red nucleus and in the (left) Pre-subgenual Frontal Cortex.

Inferred seeds are not simply a reflection of observed atrophy

As shown in **Table 2** and **3**, the inferred group average seed regions are not the same as those with the highest levels of atrophy. This confirms that early site of PD progression are not always the areas that experience the most atrophy. It also confirms that the proposed algorithm is not trivially capturing observed atrophy, but is in fact imposing additional criteria, involving seeding sparsity and the condition that the ongoing progression occur on the anatomic network.

Involvement of the striatum in PD seeding

The group average inferred seeding (**Figure 1**, **Table 1**) gives prominence to striatal structures such as putamen and accumbens. Although synuclein deposits in Lewy bodies are not quite as common in the striatum as in SNpc and the brainstem, these regions are some of the most highly affected in PD and related Lewy pathologies (Huot and Parent, 2007; Hanganu *et al.*, 2014) The fact that our inverse NDM approach was successfully able to recapitulate the key early role of these regions at the group level both confirms these regions' prominent role in disease progression, and also provides a level of validation to the proposed inverse NDM algorithm itself. This is important, since it is currently not possible to validate our results on individual patients' seeding patterns, as these data are in vivo, with no neuropathological confirmation or postmortem data.

Since both the empirical atrophy and our inferred seeding patterns suggest a strong striatal involvement, hence an important question is, *why is measured striatal Lewy pathology so low?* In PD and dementia with Lewy bodies (DLB) there has been only limited examination of striatal pathology (Jellinger and Attems, 2006). There are several possible explanations. Recent studies reveal significant striatal involvement in synucleopathies as follow. The original Braak staging of PD did not report Lewy pathology in the neostriatum (Braak et al., 2003), but in a follow up they reported the presence of neostriatal lesion in stage VI of PD (Braak et al., 2006). Mild striatal AS burden was associated with Braak stage 3, and very mild striatal AS lesions were seen in PD brains scoring Braak stages 3–5. Of PD with dementia (PDD) cases, 29% had positive synuclein lesions in the striatum. Mild to moderate synuclein pathology in caudate nucleus and putamen was seen in 76.5% of DLB brains. Almost 60% of DLB cases with initial PD phenotype (mean PD stage, 5.2) had some striatal synuclein lesions than PDD with mean PD stage, 4.0 (Jellinger and Attems, 2006). Numerous Lewy inclusions were reported in the putamen and neostriatum in DLB patients (Saito et al., 2003). A strong correlation was found between PD stages and neostriatal inclusions (Mori et al., 2008), who further reported that AS accumulates in the neostriatum at stage III initially. It is also possible that oligomeric or soluble synuclein might be more abundant than Lewy deposits in the striatum. Using novel monoclonal antibodies raised against altered synuclein, (Duda et al., 2002) uncovered an extensive burden of synuclein pathology in the striatum of LB disorders, the highest density of striatal lesions being observed in patients with DLB or a combination of AD and DLB.

Individualized seeding patterns are better predictors of future atrophy than single canonical seed regions

We find that a common incipient state of neurodegeneration cannot explain the intersubject variability observed empirically. As shown in **Figure 3**, a single-seed vector with only one non-zero entry placed either in the Amygdala or in the Substantia Nigra lead to predicted atrophy patterns that are poorly correlated with the observed data. Thus, our methods demonstrate that the early stages of the disease may exhibit incipient atrophy patterns that are more complex and variable than generally appreciated. This indicates a high level of etiologic heterogeneity in individual subjects, which makes it unlikely for any single brain region to be the source of PD pathology ramification in all subjects.

Sources of etiologic heterogeneity in PD

Our most important finding is that the estimated incipient atrophy patterns exhibit a high degree of variability across subjects, perhaps more than generally appreciated (Del Tredici and Braak, 2016). The

hierarchical clustering analysis revealed an interesting subgroup structure in the seeding patterns that were not obvious from the group seeding pattern of **Table 2**. In fact, we found two very distinct subgroups (S1 and S2, see **Figs. 4-5** and **Table 3**) from the individual subjects' inferred seeding patterns. S1 is very clearly characterized by having the Red Nucleus and Substantia Nigra as their top seed locations (all values $\geq .30$) while S2 have the Putamen, Precentral gyrus (left), and Lingual gyrus (all values $\geq .21$).

The most distinguishable signature for the S1 subgroup was the presence of the Thalamus (right/left) and the Lingual gyrus (right/left) for S2. Contrary to the name, this region has little to do with speech, and instead, is linked to processing vision (especially related to letters), logical conditions, and encoding visual memories. Our analysis show that its role in early stages of PD might be more important than generally appreciated.

The hierarchical clustering tree was also applied to the observed atrophy patterns, but the difference between the two major subgroups (A1 and A2) was not as distinct as the difference between the subgroups for seeds. We conjecture that the patients in S1 (cyan) should have worse motor symptoms due to the seeding in Substantia Nigra, Pallidum, Subthalamic Nucleus and Thalamus while patients in S2 would have difficulty in attention due to seeding involvement in lingual gyrus and frontal cortex.

In summary, we report a clear distinction in seeding patterns between the two subgroups S1 and S2 – one characterized by predominant involvement of brainstem and ventral nuclei, and the other by more widespread frontal and striatal cortices. See **Tables 3-5**. This might be indicative of two distinct etiological mechanisms operative in PD. To our knowledge, this is the first time such a clear brain related substructure has been shown in the PD cohort. Importantly, this distinction is not apparent in the individual subjects' atrophy patterns, which do not give a clear separation between these subgroups (labeled AFS1 and AFS2 in Figure 6), nor from a separate clustering directly from atrophy (labeled A1 and A2 in Figure 6). This suggests that while incipient seeding patterns show heterogeneity and clear sub-groups, these distinctions might be lost by the time the patients' atrophy patterns were measured. We believe these data point to a potentially new and unreported process of progressive convergence of PD patients, starting from divergent seeding patterns.

A

B

C

D

Fig. 6: PCA for atrophy and seed patterns. Principal Component Analysis (PCA) for (A) atrophy and seed patterns. We form the full-data atrophy matrix $X_{atrophy}$ (for plots on top) and the full-data seed matrix X_{seed} (for plots on bottom) from the 117 healthy controls (in blue) and the 232 PD patients (in red). The SVD decomposition allows us to project the data along the two first principal components (PC1 and PC2, that varies from top to bottom). We use distinct colors to mark in the mixed population the top/bottom (B) age quartiles (cyan/yellow), (C) MoCA scores (magenta/green), and (D) UPDRS3 scores (brown/beige). While there seems to be less overlap between the extreme quartiles in the bottom plots, there are no distinct clusters in any of them, which shows that the high variability of atrophy patterns within the different subjects translate also into heterogeneous seed patterns.

The challenges of validation

While we successfully infer likely sites of disease initiation from a numerical and computational point of view, our results inherit other limitations from such as the lack of a ‘gold standard’ for neuropathological validation since our analysis are based on live patients at various stages of post-onset progression (Torok *et al.*, 2018). We instead attempt to demonstrate whether the seeds are plausible based on: (i) agreement with known pathology trends in early PD; (ii) how well they are able to predict the atrophy pattern; and (iii) if they contain latent substructure that agrees with other neuropathological markers.

The regions with the highest seed values in **Table 2**, largely located within the midbrain, show strong agreement with accepted structures that are affected earlier rather than later in AS pathology in PD (Braak

et al., 2003). While our seeds do not contain the more caudal structures implicated in the earliest stages of the Braak scheme, there is disagreement about the extent to which detectable α -synucleinopathy in those regions is necessary or sufficient for pre-determining rostral manifestation of PD (Burke *et al.*, 2008). In any case, since significant atrophy changes are known to occur in the amygdala (Harding *et al.*, 2002) and in substructures of the basal ganglia such as the substantia nigra (Davie, 2008), we can determine if our individual inferred seeds are more predictive of future atrophy patterns than a single seed placed at these locations. We demonstrate that the NDM predicts patient atrophy patterns using individual seeds better than a consensus seed placed either at the SN or the amygdala (Figure 3), similar to the results in Torok *et al.*, 2018 for AD; in that study, it was also demonstrated that using a consensus seed containing 5 or 10 regions also resulted in poor predictions, indicating that this effect is independent of seed sparsity. In contrast to that study, however, we did detect two distinct subpopulations of patients with our seeds, one of which indicated strong midbrain involvement and the other more cortical area involvement (Figure 5, Table 3). The latter subpopulation is surprising in light of the generally accepted notion that the cortex harbors AS pathology only late in disease, but we note that the resultant predictions of atrophy under the NDM from the seeds of these two divergent populations are remarkably similar ($R = 0.55$ for AFS1 vs AFS2 in contrast to $R = -0.12$ for S1 vs S2). The complicated relationship between atrophy, AS pathology, and other clinical markers of PD, which limits the interpretation of these two seed subpopulations, is discussed further below

Other limitations

It is important to remark that both incipient/observed atrophy patterns are still a long way off from more tangible clinical markers such as the Montreal Cognitive Assessment (MoCA) or the Unified Parkinson's Disease Rating Scale 3 (UPDR3). In fact, Figure 6 shows a Principal Component Analysis (PCA) for both atrophy and seed patterns. We use distinct colors to mark in the mixed population the top/bottom (B) age quartiles (cyan/yellow), (C) MoCA scores (magenta/green), and (D) UPDRS3 scores (brown/beige). While there seems to be less overlap between the extreme quartiles in the bottom plots, there are no distinct clusters in any of them, which shows that the high variability of atrophy patterns within the different subjects translate also into heterogeneous seed patterns. Figure 7 shows analogous plots for data matrices restricted to PD patients alone (excluding controls). As a consequence, it remains a challenge to relate discrepancies in regional brain volumes to cognitive dysfunctions, in particular to all the different domains affected during PD such as attention and concentration, executive functions, memory, language, visuo-constructional skills, conceptual thinking, calculations, and orientation.

Fig. 7: PCA for atrophy and seed patterns. Principal Component Analysis (PCA) for (A) atrophy and seed patterns. We form the atrophy matrix $X_{atrophy}$ (for plots on top) and the seed matrix X_{seed} (for plots on bottom) restricted to PD patients alone (excluding controls). The SVD decomposition allows us to project the data along the two first principal components (PC1 and PC2, that varies from top to bottom). We use distinct colors to mark in the mixed population the top/bottom (B) age quartiles (cyan/yellow), (C) MoCA scores (magenta/green), and (D) UPDRS3 scores (brown/beige). While there seems to be less overlap between the extreme quartiles in the bottom plots, there are no distinct clusters in any of them, which shows the high variability of atrophy patterns within the different subjects translate also into heterogeneous seed patterns.

References

Ahlskog JE. Challenging conventional wisdom: The etiologic role of dopamine oxidative stress in Parkinson's disease. *Mov. Disord.* 2005; 20: 271–282.

Amunts K, Lepage C, Borgeat L, Mohlberg H, Dickscheid T, Rousseau M-E, et al. BigBrain: An Ultrahigh-Resolution 3D Human Brain Model. *Science* (80-.). 2013; 340: 1472–1475.

Braak H, Bohl JR, Müller CM, Rüb U, de Vos RAI, Del Tredici K. Stanley Fahn Lecture 2005: The staging procedure for the inclusion body pathology associated with

sporadic Parkinson's disease reconsidered. *Mov. Disord.* 2006; 21: 2042–2051.

Braak H, Braak E. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol.* 1991; 82: 239–59.

Braak H, Del Tredici K, Rüb U, de Vos RAI, Jansen Steur ENH, Braak E. Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol. Aging* 2003; 24: 197–211.

Burke RE, Dauer WT, Vonsattel JPG. A critical evaluation

- of the Braak staging scheme for Parkinson's disease. *Ann. Neurol.* 2008; 64: 485–91.
- Carbonell F, Iturria-Medina Y, Evans AC. Mathematical Modeling of Protein Misfolding Mechanisms in Neurological Diseases: A Historical Overview. *Front. Neurol.* 2018; 9: 37.
- Davie CA. A review of Parkinson's disease. *Br. Med. Bull.* 2008; 86: 109–127.
- DeLong MR, Wichmann T. Circuits and circuit disorders of the basal ganglia. *Arch. Neurol.* 2007; 64: 20–24.
- Duda JE, Giasson BI, Mabon ME, Lee VM-Y, Trojanowski JQ. Novel antibodies to synuclein show abundant striatal pathology in Lewy body diseases. *Ann. Neurol.* 2002; 52: 205–10.
- Duvernoy HM. *The Human Brain Stem and Cerebellum*. Vienna: Springer Vienna; 1995.
- Hammers A, Allom R, Koeppe MJ, Free SL, Myers R, Lemieux L, et al. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Hum. Brain Mapp.* 2003; 19: 224–47.
- Hanganu A, Bedetti C, Degroot C, Mejia-Constain B, Lafontaine A-L, Soland V, et al. Mild cognitive impairment is linked with faster rate of cortical thinning in patients with Parkinson's disease longitudinally. *Brain* 2014; 137: 1120–1129.
- Harding AJ, Stimson E, Henderson JM, Halliday GM. Clinical correlates of selective pathology in the amygdala of patients with Parkinson's disease. *Brain* 2002; 125: 2431–2445.
- Huot P, Parent A. Dopaminergic neurons intrinsic to the striatum. *J. Neurochem.* 2007; 101: 1441–1447.
- Iturria-Medina Y. Anatomical brain networks on the prediction of abnormal brain States. *Brain Connect.* 2013; 3: 1–21.
- Iturria-Medina Y, Canales-Rodríguez EJ, Melie-García L, Valdés-Hernández PA, Martínez-Montes E, Alemán-Gómez Y, et al. Characterizing brain anatomical connections using diffusion weighted MRI and graph theory. *Neuroimage* 2007; 36: 645–660.
- Iturria-Medina Y, Sotero RC, Toussaint PJ, Evans AC. Epidemic Spreading Model to Characterize Misfolded Proteins Propagation in Aging and Associated Neurodegenerative Disorders. *PLoS Comput. Biol.* 2014; 10: e1003956.
- Jellinger KA. Alpha-synuclein pathology in Parkinson's and Alzheimer's disease brain: incidence and topographic distribution--a pilot study. *Acta Neuropathol.* 2003; 106: 191–201.
- Jellinger KA, Attems J. Does striatal pathology distinguish Parkinson disease with dementia and dementia with Lewy bodies? *Acta Neuropathol.* 2006; 112: 253–60.
- Keuken MC, Bazin P-L, Crown L, Hootsmans J, Laufer A, Müller-Axt C, et al. Quantifying inter-individual anatomical variability in the subcortex using 7T structural MRI. *Neuroimage* 2014; 94: 40–46.
- Lang AE, Obeso JA. Challenges in Parkinson's disease: Restoration of the nigrostriatal dopamine system is not enough. *Lancet Neurol.* 2004; 3: 309–316.
- Leentjens AFG, Van den Akker M, Metsemakers JFM, Lousberg R, Verhey FRJ. Higher incidence of depression preceding the onset of Parkinson's disease: A register study. *Mov. Disord.* 2003; 18: 414–418.
- Lewis MM, Du G, Lee E-Y, Nasrallah Z, Sterling NW, Zhang L, et al. The pattern of gray matter atrophy in Parkinson's disease differs in cortical and subcortical regions. *J. Neurol.* 2016; 263: 68–75.
- Luk KC, Kehm V, Carroll J, Zhang B, O'Brien P, Trojanowski JQ, et al. Pathological α -Synuclein Transmission Initiates Parkinson-like Neurodegeneration in Nontransgenic Mice. *Science* (80-.). 2012; 338: 949–953.
- Masuda-Suzukake M, Nonaka T, Hosokawa M, Oikawa T, Arai T, Akiyama H, et al. Prion-like spreading of pathological α -synuclein in brain. *Brain* 2013; 136: 1128–1138.
- Mori F, Tanji K, Zhang H, Kakita A, Takahashi H, Wakabayashi K. α -Synuclein pathology in the neostriatum in Parkinson's disease. *Acta Neuropathol.* 2008; 115: 453–459.
- Pandya S, Mezas C, Raj A. Predictive Model of Spread of Progressive Supranuclear Palsy Using Directional Network Diffusion. *Front. Neurol.* 2017; 8
- Pandya S, Zeighami Y, Freeze B, Dadar M, Collins DL, Dagher A, et al. Predictive model of spread of Parkinson's pathology using network diffusion. *Submiss.* 2018
- Poewe WH, Wenning GK. The natural history of Parkinson's disease. *Ann. Neurol.* 2006; 44: VII2-16.
- Raj A, Kuceyeski A, Weiner M. A network diffusion model of disease progression in dementia. *Neuron* 2012; 73: 1204–15.

- Raj A, LoCastro E, Kuceyeski A, Tosun D, Relkin N, Weiner M. Network diffusion model of progression predicts longitudinal patterns of atrophy and metabolism in Alzheimer's Disease. *Cell Rep.* 2015; in print: 359–369.
- Raj A, Powell F. Models of Network Spread and Network Degeneration in Brain Disorders. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 2018
- Ramírez-Ruiz B, Martí MJ, Tolosa E, Bartrés-Faz D, Summerfield C, Salgado-Pineda P, et al. Longitudinal evaluation of cerebral morphological changes in Parkinson's disease with and without dementia. *J. Neurol.* 2005; 252: 1345–52.
- Rey NL, Steiner JA, Maroof N, Luk KC, Madaj Z, Trojanowski JQ, et al. Widespread transneuronal propagation of α -synucleinopathy triggered in olfactory bulb mimics prodromal Parkinson's disease. *J. Exp. Med.* 2016; 213: 1759–1778.
- Rietdijk CD, Perez-Pardo P, Garssen J, van Wezel RJA, Kraneveld AD. Exploring Braak's Hypothesis of Parkinson's Disease. *Front. Neurol.* 2017; 8: 37.
- Ross GW, Abbott RD, Petrovitch H, Tanner CM, Davis DG, Nelson J, et al. Association of olfactory dysfunction with incidental Lewy bodies. *Mov. Disord.* 2006; 21: 2062–2067.
- Ross GW, Petrovitch H, Abbott RD, Tanner CM, Popper J, Masaki K, et al. Association of olfactory dysfunction with risk for future Parkinson's disease. *Ann. Neurol.* 2008; 63: 167–173.
- Sacino AN, Thomas MA, Ceballos-Diaz C, Cruz PE, Rosario AM, Lewis J, et al. Conformational templating of α -synuclein aggregates in neuronal-glia cultures. *Mol. Neurodegener.* 2013; 8: 17.
- Saito Y, Kawashima A, Ruberu NN, Fujiwara H, Koyama S, Sawabe M, et al. Accumulation of phosphorylated alpha-synuclein in aging human brain. *J. Neuropathol. Exp. Neurol.* 2003; 62: 644–54.
- Schuurman AG, van den Akker M, Ensink KTJL, Metsemakers JFM, Knottnerus JA, Leentjens AFG, et al. Increased risk of Parkinson's disease after depression: a retrospective cohort study. *Neurology* 2002; 58: 1501–4.
- Shiba M, Bower JH, Maraganore DM, McDonnell SK, Peterson BJ, Ahlskog JE, et al. Anxiety disorders and depressive disorders preceding Parkinson's disease: a case-control study. *Mov. Disord.* 2000; 15: 669–77.
- Torok J, Maia PD, Powell F, Pandya S, Raj A. A method for inferring regional origins of neurodegeneration. *Brain* 2018; 141: 863–876.
- Del Tredici K, Braak H. Review: Sporadic Parkinson's disease: development and distribution of α -synuclein pathology. *Neuropathol. Appl. Neurobiol.* 2016; 42: 33–50.
- Del Tredici K, Rüb U, de Vos RAI, Bohl JRE, Braak H. Where Does Parkinson Disease Pathology Begin in the Brain? *J. Neuropathol. Exp. Neurol.* 2002; 61: 413–426.
- Wood SJ, Wypych J, Steavenson S, Louis JC, Citron M, Biere AL. alpha-synuclein fibrillogenesis is nucleation-dependent. Implications for the pathogenesis of Parkinson's disease. *J. Biol. Chem.* 1999; 274: 19509–12.
- Yau Y, Zeighami Y, Baker TE, Larcher K, Vainik U, Dadar M, et al. Network connectivity determines cortical thinning in early Parkinson's disease progression. *Nat. Commun.* 2018; 9: 12.
- Yonetani M, Nonaka T, Masuda M, Inukai Y, Oikawa T, Hisanaga S-I, et al. Conversion of wild-type alpha-synuclein into mutant-type fibrils and its propagation in the presence of A30P mutant. *J. Biol. Chem.* 2009; 284: 7940–50.
- Zeighami Y, Ulla M, Iturria-Medina Y, Dadar M, Zhang Y, Larcher KM-H, et al. Network structure of brain atrophy in de novo Parkinson's disease. *Elife* 2015; 4
- Zhou J, Gennatas ED, Kramer JH, Miller BL, Seeley WW. Predicting regional neurodegeneration from the healthy brain functional connectome. *Neuron* 2012; 73: 1216–27.