Exploring the reproducibility of functional connectivity alterations in Parkinson's disease

Liviu Badea[1]*, Mihaela Onu[2,3], Tao Wu[4,5], Adina Roceanu[6], Ovidiu Bajenaru[6,7]

1 Artificial Intelligence and Bioinformatics Group, National Institute for Research and Development in Informatics, Bucharest, Romania, **2** Medical Imaging Department, Clinical Hospital Prof. Dr. Th. Burghele, Bucharest, Romania, **3** University of Medicine and Pharmacy "Carol Davila", Biophysics Department, Bucharest, Romania, **4** Department of Neurobiology, Key Laboratory on Neurodegenerative Disorders of Ministry of Education, Beijing Institute of Geriatrics, Xuanwu Hospital, Capital Medical University, Beijing, China, **5** Beijing Key Laboratory on Parkinson's Disease, Parkinson Disease Centre of Beijing Institute for Brain Disorders, Beijing, China, **6** University Emergency Hospital Bucharest, Neurology Department, Bucharest, Romania, **7** University of Medicine and Pharmacy "Carol Davila", Department of Clinical Neurosciences, Bucharest, Romania

* badea.liviu@gmail.com

Abstract

Since anatomic MRI is presently not able to directly discern neuronal loss in Parkinson's Disease (PD), studying the associated functional connectivity (FC) changes seems a promising approach toward developing *non-invasive* and *non-radioactive* neuroimaging markers for this disease. While several groups have reported such FC changes in PD, there are also significant discrepancies between studies. Investigating the reproducibility of PD-related FC changes on independent datasets is therefore of crucial importance. We acquired resting-state fMRI scans for 43 subjects (27 patients and 16 normal controls, with 2 replicate scans per subject) and compared the observed FC changes with those obtained in two independent datasets, one made available by the PPMI consortium (91 patients, 18 controls) and a second one by the group of Tao Wu (20 patients, 20 controls). Unfortunately, PD-related functional connectivity changes turned out to be non-reproducible across datasets. This could be due to disease heterogeneity, but also to technical differences. To distinguish between the two, we devised a method to directly check for disease heterogeneity using random splits of a single dataset. Since we still observe *non-reproducibility* in a large fraction of random splits of the same dataset, we conclude that functional heterogeneity may be a dominating factor behind the lack of reproducibility of FC alterations in different rs-fMRI studies of PD. While *global* PD-related functional connectivity changes were non-reproducible across datasets, we identified a few *individual brain region pairs* with marginally consistent FC changes across all three datasets. However, training classifiers on each one of the three datasets to discriminate PD scans from controls produced only low accuracies on the remaining two test datasets. Moreover, classifiers trained and tested on random splits *of the same dataset* (which are technically homogeneous) also had low test accuracies, directly substantiating disease heterogeneity.

Yahr score; NC, Normal Controls; PD, Parkinson's Disease; PPMI, Parkinson's Progression Markers Initiative; ROI, Region of Interest; rs-fMRI, resting state functional Magnetic Resonance Imaging; SVM, Support Vector Machine classifier.

Introduction

Although Parkinson's disease (PD) is the second most common neurodegenerative disease after Alzheimer's disease, its diagnosis is still difficult, especially in the early premotor stages, as it is mainly based on clinical evidence. To date, there is still no unique standard diagnostic test for PD, despite the intense research efforts to develop accurate biomarkers based on blood tests or imaging scans. The best current objective tests for PD evaluate dopaminergic function in the basal ganglia by using various PET or SPECT radiotracers (e.g. DaTSCAN). But these tests make use of radioactive substances, are performed only in specialized imaging centers and can also be very expensive [1]. Moreover, the loss of dopaminergic nigro-striatal neurons is a delayed pathological event in the evolution of the disease, corresponding to Braak stages III-IV.

On the other hand, conventional (CT or MRI) brain scans of PD patients usually appear normal or with minor non-specific changes, so that conventional imaging techniques are only useful for ruling out other diseases that can be secondary causes of parkinsonism.

Therefore, since anatomic MRI is presently not able to directly discern (dopaminergic) neuronal loss in PD [2], studying the associated functional connectivity (FC) changes seems to be a promising approach toward developing *non-invasive* and *non-radioactive* neuroimaging markers for this disease.

While many groups have reported such FC changes in PD (see Table A in S1 File for a list of such studies), an in-depth analysis of existing literature revealed significant discrepancies between studies. Investigating the *reproducibility* of PD FC changes *on independent datasets* is therefore of crucial importance.

A comprehensive review and analysis of the literature related to resting-state fMRI studies of Parkinson's disease is out of the scope of the present paper [3–25] (Table A in S1 File; see also the review by Tahmasian et al. [26]). We only mention some important inconsistencies of reported functional connectivity changes in PD. Due to the crucial importance of the striatum in PD, we first discuss some inconsistencies involving striatal seeds [26]:

- Contrary to Hacker et al. [14], Helmich et al. [10] observed no significant difference in caudate functional connectivity in PD.

- On the other hand, contrary to the study Helmich et al. [10], Luo et al. [12] did not observe increased FC of the anterior putamen.

- In contrast to Hacker et al. [14], Luo et al. [12] did not find a FC decrease between the striatal seeds and the brainstem.

There are also inconsistencies involving non-striatal seeds. For example, Wu et al. [7] found disrupted FC between the pre-SMA and the left putamen, as opposed to Helmich et al. [10], who did not find a decreased FC between the putamen and pre-SMA in PD.

Since the motor symptoms are the most striking clinical manifestations in PD, many rs-fMRI studies of PD concentrate on the sensorimotor system, including the basal ganglia, while disregarding any other FC changes. On the other hand, other more unbiased studies tried to determine a more global picture of the FC changes in PD. Some even tried to develop classifiers for the disease based on rs-fMRI data [4,5,18,23], but most studies were not validated on independent datasets.

There are also some gross discrepancies involving even the sign of the main FC changes in PD. For example, Luo et al. [12] found only decreased FC in early stage PD, whereas most studies also find FC increases.

The general picture one gets from the literature is complex and at times somewhat confusing due to the numerous inconsistencies. Of course, these inconsistencies could be due to the

different disease stages analyzed, to the inherent functional heterogeneity of the disease, but also to technical differences, or to the differences in the complex data (pre)processing work-flows. Therefore, it is crucial to use the same data processing workflow to check the reproducibility of PD-related FC changes on as many independent datasets as possible.

In this work, we report a comparison between three different datasets obtained by completely independent research groups. More precisely, we acquired resting-state scans for 43 Romanian subjects (27 patients and 16 normal controls, with 2 replicate scans per subject) and compared the observed functional connectivity changes with those obtained in two independent datasets, one made available by the PPMI consortium in the US (91 patients, 18 controls) and a second one by the group of Tao Wu in China (20 patients and 20 normal controls).

This is the first study investigating the reproducibility of functional connectivity changes in Parkinson's disease on more than two datasets. Given the paucity of publicly available rs-fMRI PD datasets, we advocate the critical importance of data sharing for enabling the discovery of reproducible rs-fMRI biomarkers of PD. All the data from the present study are publicly available at FCP/INDI (http://fcon_1000.projects.nitrc.org/indi/retro/parkinsons.html).

Materials and methods

We briefly describe our approach in Fig 1A to better guide the reader through the remainder of the paper. We started by comparing the PD-related global functional connectivity changes in the 3 datasets and found them to be non-reproducible. Of course, this could be due to disease heterogeneity, but also to technical differences. To better distinguish between these two possibilities, we present a simple method to check for disease heterogeneity using random splits of a single dataset. On the other hand, we search for *individual brain region pairs* with consistent connectivity changes across all three datasets. Finally, to more directly discriminate PD scans from controls, we train multivariate machine learning classifiers on one dataset and test them on the remaining two. We also train and test classifiers on technically homogeneous random splits of the same dataset, to more directly check for disease heterogeneity.

Datasets

Three resting-state fMRI datasets of Parkinson's disease were compared in this study (see also Table 1 and the Supporting Information for more details):

- the NEUROCON rs-fMRI study of 27 PD patients and 16 normal controls (with 2 replicate scans per subject) of the Neurology Department of the University Emergency Hospital Bucharest (Romania),

- a dataset of 20 PD patients and 20 normal controls provided by the group of Tao Wu (China),

- a publicly available dataset of 91 PD patients and 18 controls of the Parkinson's Progression Markers Initiative (PPMI) study in the US.

The datasets are somewhat similar, except for PPMI, which involved patients with a diagnosis of PD for two years or less and who are not taking PD medications, while most patients from the other two studies have been under treatment (most under levodopa). Also, PPMI patients were scanned in the 'eyes open' condition. Still, we argue that our findings were not affected by these differences. Since the datasets were compared in a pairwise manner, any putative discrepancies due to the shorter disease durations in the PPMI dataset would only show

Fig 1. Overview. (A) Main steps of the analysis. (B) Using random splits of a dataset with replicate scans to check for disease (group) heterogeneity: (right) by placing different subjects (with all their replicate scans) in the two splits ("split subjects") and respectively (left) by splitting the replicates of the same subjects in the two splits ("split replicates"). (a,b,c,... correspond to subjects, while, for instance, a' and a" are replicate scans for subject a).

https://doi.org/10.1371/journal.pone.0188196.g001

up in the NEUROCON-PPMI and Tao Wu-PPMI comparisons, but not in the NEUROCON-Tao Wu comparison. This was not observed in reality.

The NEUROCON study has been approved by the University Emergency Hospital Bucharest ethics committee in accordance with the ethical standards of the 1964 Declaration of

Table 1. The 3 PD datasets compared in the present study (PD = Parkinson's disease, NC = normal controls).

Dataset	PD subjects	NC subjects	PD scans	NC scans	age PD mean (SD)	age NC mean (SD)	p(age NC-PD)	H&Y mean (SD)	disease duration mean (SD)	Scanner	Voxel(mm) TR(s)	Scan time (min) (#vols)
NEUROCON	27 (16 M)	16 (5 M)	54	31	68.7 (10.6)	67.6 (11.9)	0.76	1.92 (0.33)	4.6 (6.5)	Siemens Avanto 1.5T	3.8x3.8x5 3.48	8.05 (137)
Tao Wu	20 (11 M)	20 (12 M)	20	20	65.2 (4.4)	64.8 (5.6)	0.78	1.88 (0.63)	5.4 (3.9)	Siemens Magnetom 3T	4x4x5 2	8 (239)
PPMI	91 (63 M)	18 (14 M)	134	19	61.3 (10.2)	64.7 (9.7)	0.17	1.72 (0.48)	1.9 (1.0)	Siemens Tim Trio 3T	3.3x3.3x3.3 2.4	8.4 (210)

https://doi.org/10.1371/journal.pone.0188196.t001

Helsinki and its later amendments. All patients gave their written informed consent to participate in the study.

Preprocessing.   All datasets were preprocessed in a uniform manner. The raw scanner data in DICOM format was converted to NIfTI using dcm2nii (https://www.nitrc.org/projects/dcm2nii/) and further preprocessed using FSL (FMRIB Software Library v5 http://fsl.fmrib.ox.ac.uk/fsl/) as follows: motion correction using MCFLIRT, brain extraction with BET, spatial smoothing (Gaussian kernel FWHM 5mm) and denoising using nonlinear filtering (SUSAN), temporal high-pass filtering (with a cutoff frequency of 0.01 Hz), registration to the standard Montreal Neurological Institute MNI152 template via the anatomical T1 image (more precisely, BBR registration of the BOLD image to the T1 image, followed by 12 DOF linear+nonlinear registration of the latter to the 2mm MNI template). Nonlinear registration was performed at a resampling resolution of 4mm.

Besides the above 'standard' preprocessing workflow, we also considered alternative workflows involving global signal regression (GS) and respectively a temporal bandpass filter (0.01–0.1Hz—an ideal low-pass 0.1Hz filter was used in addition to the default FSL 0.01Hz highpass filter).

Since subject motion in the scanner has been observed to have significant influence on the functional connectivity computed from rs-fMRI data, despite motion correction (e.g. [27]), we also considered subsets of scans with low in-scanner motion (marked by the suffix '0', e.g. 'NC0' and 'PD0'–see also Table B in S1 File).

PD-related functional connectivity changes

Functional connectivity [28] is a rather loosely defined term, which encompasses many different methods used to reveal temporal correlations of BOLD activity across the brain. The simplest method consists in computing the correlations between all pairs of regions of a given brain parcellation, but more sophisticated data decomposition methods, such as Independent Component Analysis (ICA) are also widely used. Such data decomposition methods do not assume a given brain parcellation, but instead construct spatial maps grouping voxels with highly correlated timecourses. (Still, instead of being given a parcellation, such methods need a target number of components.)

In our study of the reproducibility of functional connectivity changes in PD, we used *brain parcellations constructed independently of the datasets under comparison*, rather than applying data decomposition methods, such as group-ICA, since the latter would be inherently biased towards the "training dataset". Group-ICA may obtain a better functional parcellation *for the training dataset*, but that parcellation would be less appropriate for any other independent dataset ("overfitting"), thereby introducing a bias in the analysis. To avoid these problems, we have chosen to use brain parcellations constructed independently of the datasets under comparison, including functional brain parcellations obtained by group-ICA on completely independent sets of subjects (such as the 'Stanford' functional parcellation [29]).

Moreover, to compensate for potential biases of any specific parcellation, we extended our analyses to a number of 13 different parcellations employed in other rs-fMRI studies, two anatomical (AAL, Talairach) and 11 functional (see Table 2 for more details).

For each parcellation, we computed average timecourses for each region of interest (ROI) and the *resting state functional connectivity* between each pair of ROIs (ROI_1, ROI_2) as the Fisher z-transform of the temporal correlation between the corresponding ROI timecourses:

$$FC(ROI_1, ROI_2) = z(corr(ROI_1, ROI_2)).$$

Table 2. The brain parcellations used in the functional connectivity comparisons.

Parcellation	Reference	Number of regions	Comments
AAL	Tzourio-Mazoyer et al. [30]	116	anatomic atlas
Craddock 130	Craddock et al. [31]	130	
Craddock 260		260	
Craddock 500		500	
Craddock 950		950	
Shen 100	Shen et al. [32]	93	
Shen 200		183	
Shen 300		278	
OASIS	Marcus et al. [33]	97	
Power	Power et al. [34]	264	spherical regions with a 10mm radius
Gordon_surface	Gordon et al. [35]	333	
Talairach	Talairach and Tournoux [36]	695	anatomic atlas
Stanford	Shirer et al. [29]	90	functional parcellation obtained by group-ICA

For each dataset, we determined significant PD-related FC changes by applying two-sample t-tests (with unequal sample sizes and unequal variances) to the functional connectivities of all ROI pairs. ROI-pairs with significant group differences (NC versus PD) represent regions whose functional connectivity was found to be significantly different in PD patients *in that particular dataset*. The main aim of this study is to determine whether these changes are reproducible across datasets, to enable the development of functional imaging biomarkers for PD.

Reproducibility of global functional connectivity changes in PD

Comparison of 3 different PD datasets.   We first compared the *global PD-related functional connectivity changes* across the three independent datasets NEUROCON, Tao Wu and PPMI to check to what extent these changes are reproducible. More precisely, we performed pairwise comparisons for all dataset pairs as follows.

For each pair of datasets (*i,j*) and a fixed parcellation, we checked the extent to which the *PD-related FC changes* in one dataset are correlated to the changes in the second dataset.

PD-related FC changes were quantified using t-values $t(\text{ROI}_k,\text{ROI}_l)$ from group comparisons (unpaired two-sample t-tests between NC and PD) of the functional connectivities between pairs of regions of interest $FC(\text{ROI}_k,\text{ROI}_l)$.

Then the *reproducibility R_{ij} across the two datasets i and j* was determined as the correlation between the corresponding *t*-values (viewed as a vector over all ROI pairs) for the two datasets:)

$$R_{ij} = corr(T_i, T_j) \tag{1}$$

where

$$T_i = (t_i(ROI_1, ROI_2), t_i(ROI_1, ROI_3), t_i(ROI_1, ROI_4), \ldots)$$

with $t_i(\text{ROI}_k,\text{ROI}_l)$ the *t*-value corresponding to PD-related FC changes between ROI_k and ROI_l with respect to dataset *i* (and similarly T_j for dataset *j*).

For a more intuitive graphical depiction of reproducibility across two datasets, we also constructed the *scatter-plot of ROI-pair t-values* corresponding to group comparisons in the two datasets (see Fig 2 for an example of such a scatter-plot).
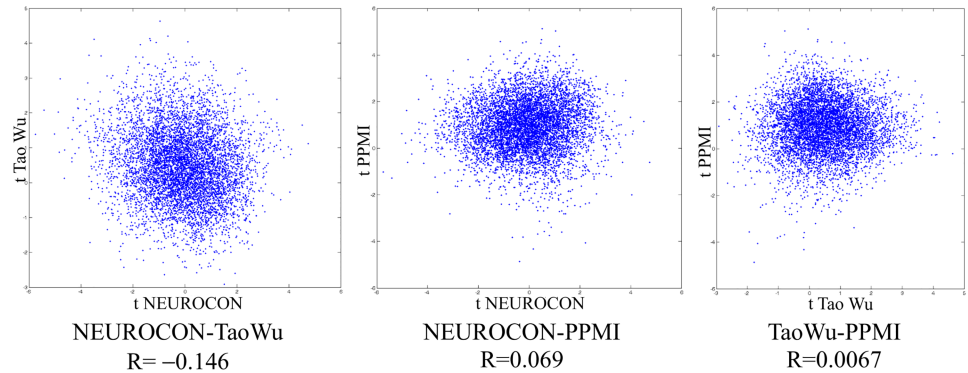
Fig 2. Scatter-plots of ROI-pair *t*-values for the three dataset pairs indicate non-reproducibility of global PD-related FC changes.

Comparing PD-related FC changes (*t*-values) in the two datasets amounts to plotting for each ROI-pair the *t*-value in dataset 1 against the *t*-value in dataset 2. We thereby obtain a scatter-plot with a point for each ROI pair. The comparison of the FC changes in the two datasets thus involves analyzing the distribution of points in the scatter-plot: ideally, perfect reproducibility would entail a diagonal distribution of points in the scatter-plot, corresponding to perfectly correlated *t*-values in the two datasets. Fig 4 depicts examples of good reproducibility, while Fig 2 shows cases of non-reproducibility across datasets.
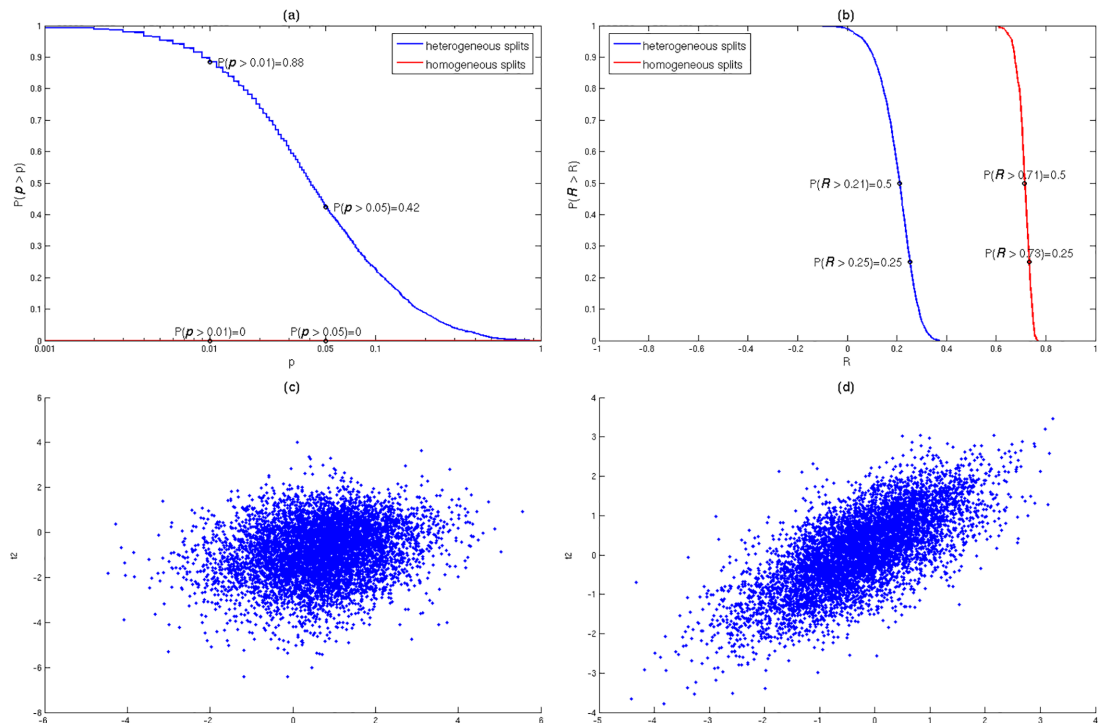
Fig 3. *Inconsistent* reproducibility of PD-related FC changes in random *heterogeneous* dataset splits and *consistent* reproducibility in random *homogeneous* dataset splits. (A) Complementary cumulative distribution function (CCDF = 1-CDF) of the reproducibility *p*-values for N_s = 2510 random *heterogeneous* splits and N_s = 325 random *homogeneous* splits. (B) CCDF of the reproducibility measure *R*. (C) A scatter-plot of ROI-pair *t*-values for a random *heterogeneous* split. (D) A scatter-plot of ROI-pair *t*-values for a random *homogeneous* split.
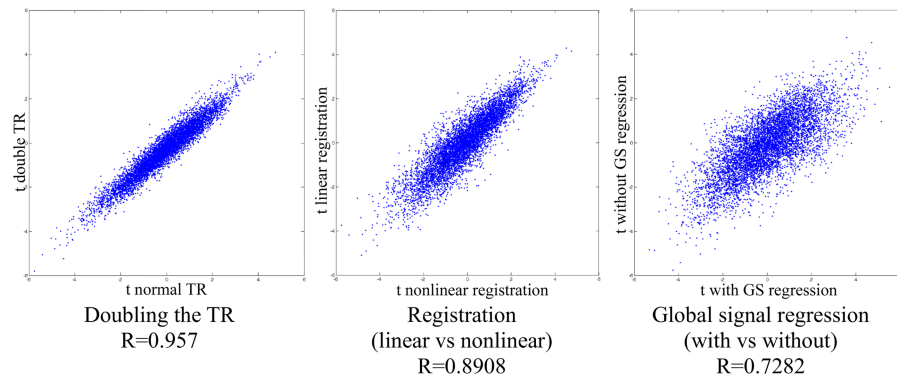
Fig 4. Reproducibility when changing various technical factors or preprocessing options.

The correlation of *t*-values for the two datasets $R_{ij} = corr(T_i, T_j)$, as introduced above in (1), can be viewed as an aggregate measure of the *reproducibility across the two datasets i* and *j*.

To obtain a more quantitative measure of the *statistical significance* of such a correlation R_{ij} between datasets, we performed permutations of the group labels (NC and PD) *independently* for the two datasets and computed the *p-value of the R_{ij} value* as the fraction of permutations σ for which the dataset correlation w.r.t. the permuted data $R_{ij}^{(\sigma)}$ exceeds the real one (R_{ij}):

$$p_{ij} = |\{permutation\ \sigma | R_{ij}^{(\sigma)} \geq R_{ij}\}|/N \tag{2}$$

where *N* is the total number of permutations. All our permutation tests involved $N = 1000$ permutations.

Various factors have been mentioned in the literature to affect functional connectivity measures:

- subject motion in the scanner [27],

- global signal regression (with or without) [37,38],

- the choice of the parcellation.

To study the influence of these factors on our reproducibility results, we also considered subsets of scans with *low in-scanner motion* (marked by the suffix '0', e.g. 'NC0' and 'PD0'), repeated our analyses with global signal regression, bandpass filtering and performed the comparisons using all 13 brain parcellations previously mentioned. Since the PPMI data has been acquired in several different imaging centers, we also considered a potentially more homogeneous subset of scans acquired in a single center (center number 32, with the largest number of PD patient and normal control scans), referred in the following by the suffix 'center32'.

Comparison of random splits of the same PD dataset. As already mentioned in the Introduction, the observed lack of reproducibility of global FC changes across datasets could be due to disease heterogeneity, but also to technical differences. To distinguish between these two possibilities, we devised a method to directly check for disease heterogeneity using random splits of a single dataset with replicate scans. Technical differences can then be excluded since all the scans have been acquired under identical technical conditions. More precisely, since in the NEUROCON study we acquired two replicate scans for each subject, we constructed two *homogeneous dataset splits* simply by using (different) scans of the same subjects. Additionally, two *heterogeneous dataset splits* can be obtained by placing different subjects

(with all their scans) in each split. In other words, instead of comparing two distinct datasets, we compared two random splits of the same dataset, either:

(a) by placing different subjects in the two splits, with all the replicate scans of a subject in the same split ("*split subjects*", *heterogeneous split*), or

(b) by placing each replicate scan of the same subject in a different split, so that the two splits contain (different) scans of the same subjects ("*split replicates*", *homogeneous split*).

Dataset splits (b) are homogeneous since they contain scans of the same subjects, while splits (a) are heterogeneous since they contain scans of different subjects. Therefore, *consistent reproducibility* across *all* random *heterogeneous splits* would indicate *disease homogeneity*, while *non-reproducibility* in a large fraction of random heterogeneous splits would imply *disease heterogeneity*. (In both cases, we expect to observe consistent reproducibility across the homogeneous splits, at least as long as the technical noise is not dominating the biological signal.) A diagram of our method is shown in Fig 1B.

As in the pairwise comparisons between different datasets, we used permutation tests and formula (2) to compute p-values of the reproducibility across split datasets, for both the heterogeneous ("split subjects") and the homogeneous ("split replicates") datasets. Due to the random nature of the splits, we repeated the analysis for $N_s > 1000$ different random splits of the original data. To assess the fraction of (non-)reproducible splits, we determined the empirical cumulative distribution function (CDF) of the reproducibility p-values for the N_s random splits.

The analysis was also repeated for the data with global signal regression.

Influence of technical factors, preprocessing and parcellation. We also studied the influence on reproducibility of certain key technical factors and preprocessing steps, such as:

- the repetition time (TR),

- linear vs. nonlinear registration,

- global signal regression,

- the specific brain parcellation used for evaluating functional connectivity.

The AAL parcellation (which is typical) was used whenever not specified otherwise.

The *repetition time* might, in principle, influence the measured low-frequency rs-fMRI fluctuations and indirectly the functional connectivities (which are temporal correlations). To study the influence of the repetition time on reproducibility, we constructed a synthetic dataset with a *double TR* by leaving out every second time-point from the NEUROCON timeseries data (for each scan and each voxel). We then analyzed with our method the reproducibility of group changes in functional connectivity between the original NEUROCON dataset and the synthetic one with a double TR.

To study the impact of *registration* on reproducibility, we preprocessed the NEUROCON data both with *linear* and *nonlinear registration* to the MNI 152 template and determined the reproducibility of group changes in functional connectivity between the two resulting datasets.

Since *global signal regression* (GSR) has been observed to be very effective at removing scanning artifacts [38], including motion artifacts, we also studied the reproducibility of FC changes between the NEUROCON dataset processed with GSR and the same dataset processed without GSR.

To avoid potential biases of any specific parcellation, we repeated our pairwise comparisons between the 3 PD datasets using all *13 different parcellations* mentioned above, including functional and anatomic parcellations, with a wide range of numbers of regions of interest (90 to 950).

Reproducibility of the individual differentiating FC changes

The reproducibility analysis performed above involves *global* functional connectivity changes, i.e changes in the FC of *all* ROI pairs, not just the ones that differentiate PD from normal controls. Even with non-reproducible global FC changes, it might be in principle possible that only a very few brain region pairs might still reproducibly differentiate PD patients from controls. To this end, we also studied *individual brain region pairs* with FC changes that are significant w.r.t. *all* datasets.

More precisely, for each ROI-pair, we compute $max(p)$, the largest (least significant) of the three p-values obtained in the three datasets (separately for the FC increases and respectively decreases) and sort the ROI-pairs in increasing order of this $max(p)$. The most significant $min(max(p))$ of these $max(p)$ corresponds to the ROI-pair with the best overall significance with respect to the 3 datasets, as all other ROI-pairs have larger (less significant) *p*-values with respect to at least one dataset.

Finally, to assess the statistical significance of such a best ROI-pair, we use a permutation test (of the disease labels in each dataset) to check the fraction of random permutations with a more significant (smaller) $min(max(p))$ than the real data (we performed $N = 1000$ random permutations).

$$p(\min(\max(p_+))) = |\{permutation \ \sigma| \leq \min(max(p_+))\}|/N$$

where $min(max(p_+))$ corresponds to FC increases in NC versus PD. A similar relation holds for the FC decreases $min(max(p_-))$.

The analysis was repeated for all 13 parcellations considered in this study ([Table 2](#)).

Learning classifiers for discriminating PD-related FC changes

We used machine learning techniques to learn classifiers that discriminate PD from controls using functional connectivities between ROI pairs as features.

First, we trained classifiers on each one of the 3 datasets (NEUROCON, Tao Wu, PPMI) and tested them on the other two datasets. Both Linear Support Vector Machines (SVM) and Gaussian Naïve Bayes (GNB) classifiers were tested, with progressively increasing numbers of features: $N = 10, 50, 100, 500, 5000$ (functional connectivities between ROI pairs). The N features selected were the best discriminating ROI pair functional connectivities, based on unpaired *t*-tests between normal and PD scans. As the two classes (NC-Normal Controls and PD-Parkinson's Disease) are not balanced in all 3 datasets, we employ the *average accuracy* $Aacc = (acc(NC)+acc(PD))/2$ for assessing the performance of the classifiers (a random classifier is expected to have an average accuracy of 0.5).

Since the different datasets are not technically homogeneous, we also trained and tested classifiers on random splits of the same dataset, to check to what extent the low accuracies are due to technical differences, or to disease heterogeneity. More precisely, we performed 10,000 random splits in half of each dataset, trained a classifier on one half and tested it on the other.

Results

ROI-pairs with significant group differences (NC versus PD) in functional connectivity were found in all three PD datasets: NEUROCON, Tao Wu and PPMI (Tables C and D in S1 File). However, these changes seemed at first sight to be *distinct in each dataset*. Our main aim in this paper has been to systematically investigate the reproducibility of the PD-related FC changes across independent validation datasets.

PD-related FC changes are non-reproducible across 3 datasets

The reproducibility of global PD-related functional connectivity changes was determined by pairwise comparisons between three independent datasets: NEUROCON, Tao Wu and respectively PPMI. Fig 2 shows the scatter-plots of ROI-pair t-values (corresponding to the group comparison NC-PD) for the three dataset pairs, indicating a lack of reproducibility of global FC changes in PD. (Perfect reproducibility would correspond to a diagonal distribution of points corresponding to ROI pairs with perfectly correlated *t*-values with respect to both datasets.) Moreover, discriminating ROI-pairs situated in the upper right and respectively lower left corners of one plot are not discriminating in the other plots.

For a more quantitative measure of the reproducibility of FC changes between two datasets, we computed the Pearson correlation between *t*-values (viewed as vectors of over all ROI pairs) with respect to each dataset (*R* values shown in Fig 2.) We also estimated the statistical significance (*p*-values) of these reproducibility measures by permutation tests of the group labels independently for the two datasets–Table 3 shows the reproducibility measure and associated p-value for various pairwise comparisons between the three datasets, with standard highpass preprocessing (>0.01Hz), global signal regression (GS) and respectively bandpass filter (0.01–0.1Hz). Since in-scanner motion may influence FC measures, we present not only a comparison between the full patient and normal control cohort, but also that corresponding to a subset of scans with low in-scanner motion (denoted by the suffix '0'). Moreover, since PPMI data were acquired at many different centers, we also considered the restriction of the PPMI data to the scans from a single center (suffix 'center32'). The AAL parcellation was used in this case, but we also study the influence of the parcellation later on.

A clear lack of reproducibility of global PD-related FC changes is observed in all the three dataset pairs. This is the first study comparing three independent rs-fMRI datasets of PD. The fact that we compare 3 datasets is very important, as it lowers the probability that the lack of reproducibility is due to a dataset that may be "faulty" in some sense—in that case, with 3 datasets we might still observe reproducibility with respect to the remaining dataset pair (which we do not see in reality).

Inconsistent reproducibility of FC changes in heterogeneous dataset splits indicates disease heterogeneity

The non-reproducibility across 3 datasets mentioned above seems to be due to disease heterogeneity, but it could also be due to technical differences. To exclude the latter

Table 3. Reproducibility measure *R* and associated *p*-value *p* for various pairwise comparisons between datasets with standard highpass preprocessing ('standard'), global signal regression (GS) and respectively bandpass filter (BP) ('0'–'low in scanner motion', 'center32'–scans performed at a single PPMI center).

Dataset 1	Group contrast 1	Dataset 2	Group contrast 2	standard		GS		BP	
				R	*p*	*R*	*p*	*R*	*p*
NEUROCON	NC-PD	Tao Wu	NC-PD	-0.145955	0.949	-0.0679698	0.804	0.000845579	0.476
	NC0-PD0		NC0-PD0	-0.0832455	0.847	0.035725	0.300	0.0132475	0.437
NEUROCON	NC-PD	PPMI	NC-PD	0.0692377	0.257	0.10039	0.138	-0.163251	0.937
	NC0-PD0		NC0-PD0	0.0401831	0.352	0.0308243	0.373	-0.0824776	0.800
	NC-PD		NC_center32-PD_center32	-0.0122797	0.526	0.0188561	0.430	-0.142598	0.914
	NC0-PD0		NC0_center32-PD0_center32	-0.0873994	0.813	-0.0894912	0.832	-0.152545	0.945
Tao Wu	NC-PD	PPMI	NC-PD	0.00673472	0.466	0.0228885	0.403	0.0948706	0.153
	NC0-PD0		NC0-PD0	-0.0182896	0.560	0.044843	0.298	0.0545817	0.293
	NC-PD		NC_center32-PD_center32	-0.0446155	0.687	-0.032841	0.666	0.0243423	0.400
	NC0-PD0		NC0_center32-PD0_center32	-0.0712155	0.812	-0.0176842	0.597	-0.0436096	0.681

possibility, we checked for disease heterogeneity using random splits of a single dataset with replicate scans (NEUROCON), all of which have been acquired under identical technical conditions.

(a) We first constructed random splits by placing different subjects in the two splits, with all the replicate scans of a subject in the same split ("*split subjects*", *heterogeneous splits*). As can be seen in Fig 3A (blue curve), a large fraction of these random splits display non-reproducible functional connectivity changes. More precisely, 88% of the random heterogeneous splits show non-reproducibility at the $p > 0.01$ level and 42% at the $p > 0.05$ level. Fig 3B shows the complementary cumulative distribution function (1-CDF) for the corresponding reproducibility measure R (blue curve), while Fig 3C presents a typical scatter-plot of ROI-pair t-values for a random heterogeneous split (one with R equal to the median).

(b) Next, we constructed random splits by placing each replicate scan of the same subject in a different split, so that the two splits contain (different) scans of the same subjects ("split replicates", *homogeneous splits*). In contrast to (a), all homogeneous splits showed reproducibility at the $p < 10^{-3}$ level (Fig 3A, red curve). Fig 3B shows the complementary CDF for the reproducibility measure (red curve)–note the significantly higher reproducibility (R) values for the homogeneous splits (red curve, median $R = 0.71$) as compared to the heterogeneous splits (blue curve, median $R = 0.21$). Fig 3D displays a typical scatter-plot of ROI-pair t-values for a random homogeneous split.

The observed *non-reproducibility* in a large fraction of *heterogeneous dataset splits* indicates *disease heterogeneity*, in line with the comparison between the 3 independent PD datasets. As a control, we did indeed observe *consistent* reproducibility with respect to *all homogeneous* dataset splits, demonstrating that the technical noise could not have been the dominating factor behind the erratic non-reproducibility in heterogeneous splits.

The fact that the well-known *clinical* heterogeneity of Parkinson's disease is also accompanied by heterogeneity in resting state *functional connectivity* may not *retrospectively* be a big surprise to an experienced neurologist, although its exact extent could not have been estimated a priori, before analyzing the data. However, does this FC heterogeneity in PD also imply the lack of practical usefulness of rs-fMRI functional connectivity? Are there any other conditions that can be reliably differentiated using resting state functional connectivity? To answer these questions, we applied our approach to a different, potentially more homogeneous contrast, namely that between eyes open and eyes closed resting state conditions in healthy volunteers. Repeating our analysis of reproducibility of FC group changes on random splits of the Beijing eyes open-eyes closed dataset [39] (see Supporting Information) revealed *reproducibility* ($p < 0.05$) not just in the *homogeneous* dataset splits, but also in the *heterogeneous* ones (Fig A in S1 File–only 6% of the heterogeneous and just 0.8% of the homogeneous random splits were non-reproducible at the $p > 0.05$ level).

Summing up our findings, from the point of view of global FC changes, *Parkinson's disease* is *heterogeneous*, as opposed to the *eyes open-eyes closed* contrast, which is much more homogeneous (Table 4).

Table 4. Summary of reproducibility of global functional connectivity changes in Parkinson's Disease and respectively the eyes open-eyes closed contrast.

Reproducible global FC changes?	PD	EO-EC
different datasets	**no**	
split subjects (heterogeneous splits)	**inconsistent** (**no** in a large fraction of splits)	yes
split replicates (homogeneous splits)	yes	yes

https://doi.org/10.1371/journal.pone.0188196.t004

Influence of technical factors and preprocessing on reproducibility

We found good reproducibility when changing various technical factors or processing options of the NEUROCON data, such as (see Fig 4):

- doubling the repetition time (TR),

- registration (linear versus nonlinear),

- global signal regression (with versus without).

This is in line with our conclusion that functional heterogeneity, rather than these technical factors, is the dominating factor behind the lack of reproducibility of FC changes in different rs-fMRI studies of Parkinson's disease.

We also tested the influence of various rs-fMRI denoising methods on the reproducibility of PD-related FC changes, such as ICA-FIX [40,41], or regression of the mean white matter and/or cerebrospinal fluid signal—none of these denoising methods changed the observed non-reproducibility (data not shown).

We also observed no improvement in reproducibility across random splits of the NEURO-CON dataset after regressing out potential confounders, such as age, gender, or disease duration (data not shown).

Influence of parcellation on reproducibility

We have argued that functional connectivity must be computed with respect to an unbiased parcellation (i.e. one that hasn't been constructed from any of the analyzed datasets). However, any given parcellation has also specific biases that may in principle affect the capacity to discriminate between PD and normal controls—especially relevant factors are the average ROI size and the number of ROIs. Testing the reproducibility of the PD-related global FC changes using 13 different parcellations, with varying numbers of ROIs (between 90 and 950, see Table 2), revealed a lack of reproducibility regardless of parcellation, or dataset pair (Table 5). (The NEUROCON-PPMI comparison was marginally significant ($p = 0.05$) for the NC-PD contrast, but this significance didn't survive perturbations such as selecting just the 'center32'

Table 5. Reproducibility measure and associated p-value for 13 parcellations and all three dataset pairs (NC-PD contrast).

Parcellation	NEUROCON-TaoWu		NEUROCON-PPMI		TaoWu-PPMI	
	R	p	R	p	R	p
Craddock130	-0.011116	0.554	0.0927652	0.191	-0.0613781	0.747
Craddock260	-0.00871389	0.560	0.0775174	0.205	-0.0770479	0.864
Craddock500	0.00488102	0.469	0.079955	0.172	-0.0546894	0.839
Craddock950	0.00952484	0.409	0.0694818	0.174	-0.0347004	0.748
Shen100	-0.0490741	0.708	0.0958863	0.187	-0.0567748	0.734
Shen200	-0.032232	0.641	0.0936098	0.172	-0.0618746	0.799
Shen300	-0.0247647	0.632	0.0829943	0.159	-0.0729234	0.871
OASIS	0.00698018	0.482	0.0205063	0.411	-0.0155159	0.598
Power264	0.0533032	0.166	0.0812484	0.141	-0.0153593	0.600
Gordon_surface	0.0381087	0.262	0.025185	0.391	-0.0435482	0.748
Talairach	-0.0273596	0.686	0.0263891	0.357	0.0598013	0.129
Stanford	-0.0600487	0.730	0.185859	0.050	-0.081526	0.819
AAL	-0.145955	0.949	0.0692377	0.257	0.00673472	0.466

https://doi.org/10.1371/journal.pone.0188196.t005

Table 6. Best ROI-pairs with FC decreases in PD (t>0, corresponding to NC>PD) and max(p) < 0.05 (for the Power264 parcellation, without global signal regression).

max(p)	ROI1-ROI2	p NEUROCON (NC-PD)	p TaoWu (NC-PD)	p PPMI (NC-PD)	t NEUROCON (NC-PD)	t TaoWu (NC-PD)	t PPMI (NC-PD)
0.012526	sphere5(-21,-31,61)-sphere5 (15,-77,31)	0.012526	0.0018915	0.011638	2.5541	3.3398	2.7402
0.016354	sphere5(-42,45,-2)-sphere5 (43,-78,-12)	0.0026347	0.016354	0.0079673	3.1172	2.5153	2.905
0.025311	sphere5(-21,-31,61)-sphere5 (-24,-91,19)	0.0019939	0.025311	0.025232	3.2025	2.3298	2.3949
0.038357	sphere5(-38,-33,17)-sphere5 (20,-86,-2)	0.038357	0.00041882	0.014314	2.1104	3.8665	2.653
0.038529	sphere5(37,-81,1)-sphere5 (38,-17,45)	0.038529	0.02727	0.038269	2.1123	2.297	2.2003
0.038634	sphere5(-21,-31,61)-sphere5 (-26,-90,3)	0.012064	0.038634	0.033842	2.5863	2.1425	2.2603
0.044406	sphere5(-21,-31,61)-sphere5 (29,-77,25)	0.0085123	0.027058	0.044406	2.7193	2.3019	2.1243
0.046861	sphere5(-21,-31,61)-sphere5 (-40,-88,-6)	0.0077414	0.046861	0.017283	2.7342	2.0557	2.55

scans from PPMI (p = 0.259), or restriction to the low motion scans NC0-PD0 (p = 0.26), or NC0-PD0_center32 (p = 0.555).)

Marginally significant individual differentiating FC changes in PD

Despite non-reproducibility of PD-related *global* FC changes across different datasets, a small number of ROI-pairs that distinguish PD from controls may nevertheless, in principle, show reproducible changes across datasets. To check for this possibility, we concentrated on *individual brain region pairs* with FC changes that are significant w.r.t. *all* datasets, by sorting the ROI-pairs according to their least significance *max(p)* with respect to *all* datasets.

For example, Table 6 shows the ROI-pairs with FC decreases in PD (i.e. positive t-values, corresponding to NC>PD) and max(p) < 0.05 for the Power264 parcellation, without global signal regression. The best ROI-pair has max(p_+) = 0.0125, so $min(max(p_+))$ = 0.0125.

To check whether this $min(max(p_+))$ is statistically significant, we performed permutation tests as described. Table 7 lists these $min(max(p_\pm))$ values as well as their associated significance

Table 7. Significance of $min(max(p_\pm))$ values for all 13 parcellations (no global signal regression).

Parcellation	$min(max(p_+))$	$min(max(p_-))$	$p(min(max(p_+)))$	$p(min(max(p_-)))$
AAL	0.0607935	0.112189	0.208	0.426
Craddock130	0.0503968	0.242712	0.152	0.831
Craddock260	0.0294203	0.148728	0.170	0.843
Craddock500	0.0202304	0.0909226	0.206	0.859
Craddock950	0.0172931	0.0542509	0.316	0.815
Shen100	0.0371887	0.376232	**0.055** * *	0.905
Shen200	0.0425589	0.151199	0.195	0.747
Shen300	0.0324906	0.115015	0.199	0.804
OASIS	0.0697703	0.3402	0.208	0.938
Power264	0.012526	0.118716	**0.033** *	0.846
Gordon_surface	0.0275321	0.11095	0.185	0.846
Talairach	0.00701726	0.0339528	**0.032** *	0.522
Stanford	0.0599084	0.281706	0.145	0.820

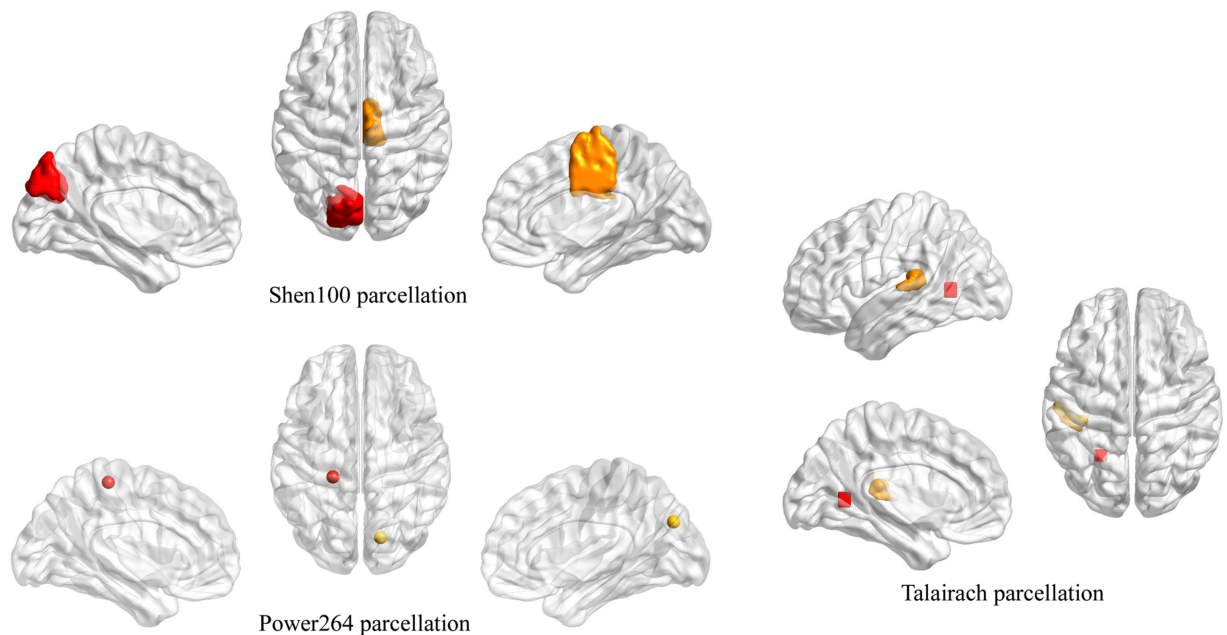**Table 8. Marginally significant FC changes w.r.t. all 3 datasets (decreased in PD).**

Parcellation	$p(min(max(p_+)))$	$min(max(p_+))$	ROI$_1$	ROI$_2$
Talairach	0.032	0.00701726	(-24,-58,4) left visual association area, lingual gyrus, BA18	(−38, −32,16) left superior temporal gyrus, BA41, planum temporale / parietal operculum
Power264	0.033	0.012526	sphere5(-21,-31,61) left postcentral / precentral gyrus	sphere5(15, −77,31) right cuneus
Shen100	0.055	0.0371887	L.BA19.3 left cuneus, precuneus	R.BA6.1 right SMA, middle cingulate

https://doi.org/10.1371/journal.pone.0188196.t008

$p(min(max(p_{\pm})))$ for all 13 parcellations. Only two out of the 13 parcellations yielded significant ROI-pairs at the p<0.05 significance level ('Power264' and 'Talairach'), while a third parcellation produced only marginally significant ROI-pairs ('Shen100', p = 0.055)—see Table 8 and Fig 5. These (marginally) significant ROI-pairs involve visual-sensorimotor, respectively visual-parietal association areas. Whether these ROI-pair changes are more widely reproducible or not will have to await the release of more publicly-available PD rs-fMRI datasets.

Testing classifiers for discriminating PD-related FC changes

Training classifiers on functional connectivity data for each one of the 3 datasets (NEURO-CON, Tao Wu, PPMI) and testing them on the other two datasets produced average accuracies on test data in the range 0.225–0.7 (mean 0.497, standard deviation 0.073), while a random classifier is expected to have an average accuracy of 0.5. Fig 6 shows the corresponding *average accuracies Aacc = (acc(NC)+acc(PD))/2* for standard preprocessing (with the default FSL high-pass filter at 0.01Hz), global signal regression and respectively bandpass filtering (0.01–0.1Hz) for both linear SVM and Gaussian Naïve Bayes (GNB) classifiers with *N* = 5000 features (out

Fig 5. Marginally significant FC changes w.r.t. all 3 datasets. The ROIs were mapped onto the brain surface using BrainNet Viewer [42] (http://www.nitrc.org/projects/bnv/).

https://doi.org/10.1371/journal.pone.0188196.g005
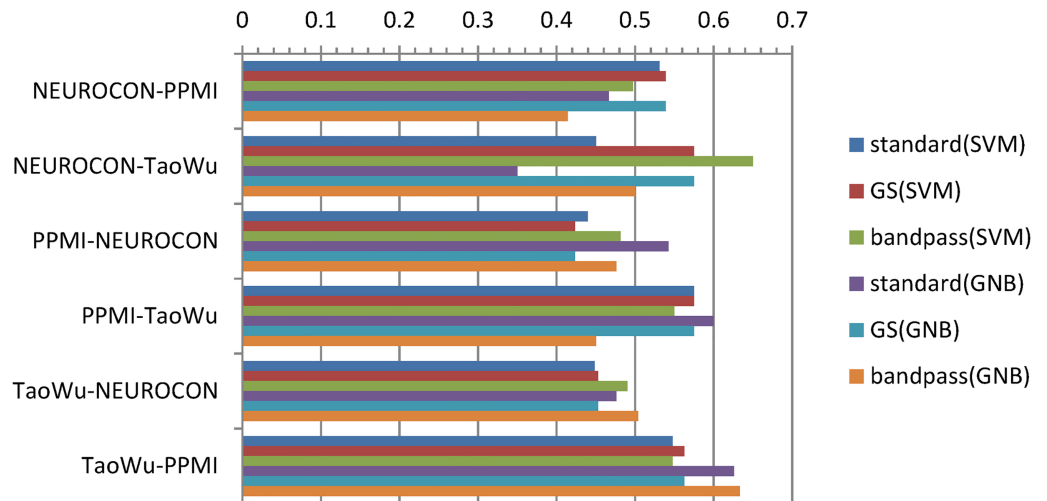
Fig 6. Average accuracies for classifiers trained on dataset 1 and tested on dataset 2 for all dataset pairs using standard preprocessing ('standard'), global signal regression (GS) and respectively bandpass filtering (0.01–0.1Hz). Here, SVM (linear Support Vector Machine) and GNB (Gaussian Naïve Bayes) classifiers used N = 5000 features—see Fig B in S1 File for classifier accuracies for varying N. As an example, NEUROCON-PPMI denotes classifiers trained on NEUROCON and tested on PPMI data.

https://doi.org/10.1371/journal.pone.0188196.g006

of the total of 6670 ROI pairs of the AAL parcellation). See Fig B and Table E in S1 File for the accuracies of classifiers with N = 10,50,100,500,5000 features.

Since for each *training* dataset (for example NEUROCON), we have two different *test* datasets (PPMI and TaoWu in our example), we also computed an *aggregated* average accuracy for each dataset by taking the mean of the two average accuracies corresponding to the two remaining test datasets (Aacc(dataset-dataset1)+Aacc(dataset-dataset2))/2. The resulting aggregated average accuracies were low, in the range 0.336–0.591 (mean 0.497, standard deviation 0.0522, compared to 0.5 for a random classifier; see also Fig 7).

Since the three datasets are not technically homogeneous, we also trained and tested classifiers on random splits of the same dataset, to check to what extent the low accuracies are due to technical differences, or to disease heterogeneity. Fig 8 shows the average accuracies for 10,000 random splits in half of each dataset and various preprocessing options. Again, the means of the average accuracies over the 10,000 tests were low, in the range 0.51–0.66, reinforcing the evidence for disease heterogeneity.

Finally, for a a more direct graphical depiction of the heterogeneity of the functional connectomes of the PD patients, we have applied *consensus NMF clustering* [43] of the PD-related FC changes for a progressively increasing number of clusters (k = 2,...,18, Fig C in S1 File). Note that besides the consistent grouping of the replicate scan pairs for each patient, it is difficult to single out an optimal number of clusters k.

Discussion

The accelerated increase in the number of functional connectivity studies of Parkinson's Disease requires a consolidation of the knowledge in this field for enabling the development of clinically relevant rs-fMRI markers for this disease. Unfortunately however, there are many inconsistencies between published works and virtually no high confidence reproducibility studies.

This is the first study investigating the reproducibility of functional connectivity changes in Parkinson's disease on more than two datasets. The fact that we use a *uniform data processing*
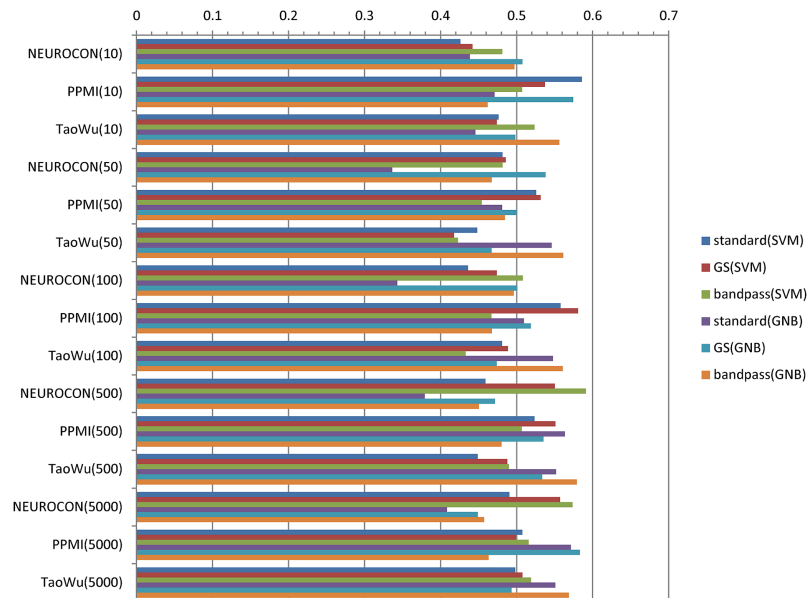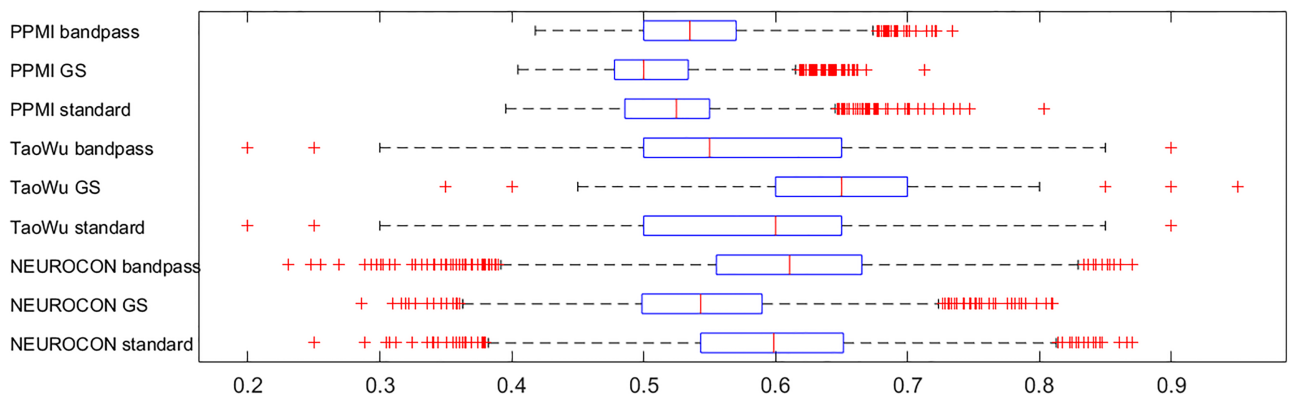
Fig 7. *Aggregated* average accuracies for classifiers trained on each of the 3 datasets using standard preprocessing ('standard'), global signal regression (GS) and respectively bandpass filtering (0.01–0.1Hz). Classifiers were trained with N = 10,50,100,500,5000 features. As an example, NEUROCON(10) refers to the aggregated accuracy (Aacc(NEUROCON-PPMI) + Aacc(NEUROCON-TaoWu))/2 for classifiers trained on NEUROCON and tested on PPMI and respectively TaoWu data using N = 10 features. SVM—linear SVM classifier, GNB—Gaussian Naïve Bayes classifier.

workflow for all datasets excludes a large number of technical factors as potential culprits for the observed differences between datasets. Also, the fact that our comparison involves *three* datasets is essential, as it lowers the probability that the observed lack of reproducibility is due to a problematic dataset—in such a case, with 3 datasets we might still observe reproducibility with respect to the remaining dataset pair, something which we do not see in reality.

To better clarify the issue, we devised a method to directly check for disease heterogeneity using random splits of a *single dataset* with replicate scans. Technical differences can then be excluded since all the scans have been acquired under identical technical conditions. The fact

Fig 8. Average accuracies for classifiers trained and tested on split data *from the same dataset* using standard preprocessing ('standard'), global signal regression (GS) and respectively bandpass filtering (0.01–0.1Hz). An SVM classifier with N = 5000 features was used.

that we still observe non-reproducibility in a significant fraction of random subsamples of each individual dataset (these subsamples being *technically homogeneous* as they come from the same dataset), suggests that *functional heterogeneity* may be a dominating factor behind the lack of reproducibility of functional connectivity alterations in different resting state fMRI studies of Parkinson's disease.

This could be due to the heterogeneous multi-lesional topography and progression of the neurodegenerative process, possibly accompanied by variable compensatory functional circuit changes, as well as by changes due to dopaminergic medication [26].

The heterogeneity of the functional connectome changes in PD is also more directly apparent in the consensus clustering plots (Fig C in S1 File).

While *global* PD-related functional connectivity differences were non-reproducible across datasets, we identified a few *individual ROI pairs* with marginally consistent FC differences across all three datasets. However, finding out whether these differences are more widely reproducible or not will have to await the release of more public PD datasets.

Additionally, we applied more sophisticated multivariate machine learning techniques to learn classifiers that discriminate PD from controls using functional connectivities between ROI pairs as features. However, training classifiers on each one of the three datasets (NEURO-CON, Tao Wu, PPMI) produced only low accuracies on the remaining two (test) datasets, in line with the preceding results. Furthermore, since the three datasets are not technically homogeneous, we also trained and tested classifiers on random splits *of the same dataset*, to more directly check to what extent the low accuracies are due to technical differences, or to disease heterogeneity. Again, we obtained low average accuracies (with means in the range 0.51–0.66), reinforcing the evidence for disease heterogeneity. Interestingly, these results are consistent with a recent study [44] on multisite generalizability of *schizophrenia* diagnosis based on functional brain connectivity, which reported multisite classification accuracies below 70%, in contrast to over 30 previously published, largely single-site schizophrenia studies, whose average reported classification accuracy exceeds 80%.

Therefore, given the paucity of publicly available rs-fMRI PD datasets, we advocate the critical importance of data sharing for enabling the discovery of reproducible and *clinically useful* functional imaging biomarkers of PD. In this regard, we view our study as an important first step towards more refined reproducibility studies that would be possible only with more publicly available datasets. In view of the many inconsistencies found in the published literature on PD-related functional connectivity changes, we strongly argue for a direct computational comparison of PD rs-fMRI datasets using a uniform data processing workflow, to avoid publication bias as well as processing workflow differences in the separate studies.

Limitations

The present study has concentrated on PD-related changes in *functional connectivity* (loosely viewed as *correlations* between different regions of interest), rather than changes in fluctuations of the *amplitude* of the rs-fMRI signal. In a complementary study, Wu et al. [45] observed PD-related changes in ALFF, but with rather limited reproducibility. An in-depth analysis of the reproducibility of PD-related differences in the *amplitude* of fluctuations is out of the scope of the present paper.

Supporting information

S1 File. Supplementary material.
(DOC)

Acknowledgments

Author Contributions

Conceptualization: Liviu Badea, Ovidiu Bajenaru.

Data curation: Liviu Badea, Mihaela Onu, Tao Wu, Adina Roceanu.

Formal analysis: Liviu Badea.

Funding acquisition: Mihaela Onu.

Investigation: Liviu Badea, Mihaela Onu, Tao Wu, Adina Roceanu, Ovidiu Bajenaru.

Methodology: Liviu Badea, Ovidiu Bajenaru.

Project administration: Mihaela Onu.

Resources: Mihaela Onu, Tao Wu, Adina Roceanu, Ovidiu Bajenaru.

Software: Liviu Badea.

Supervision: Liviu Badea, Ovidiu Bajenaru.

Validation: Liviu Badea.

Visualization: Liviu Badea.

Writing – original draft: Liviu Badea.

Writing – review & editing: Liviu Badea, Mihaela Onu, Tao Wu, Ovidiu Bajenaru.

References

1. Agrawal M, Biswas A. Molecular diagnostics of neurodegenerative disorders. Frontiers in molecular biosciences. 2015; 2.

2. Tuite PJ, Mangia S, Michaeli S. Magnetic resonance imaging (MRI) in Parkinson's disease. Journal of Alzheimer's disease & Parkinsonism. 2013 Mar 25:001.

3. Göttlich M, Münte TF, Heldmann M, Kasten M, Hagenah J, Krämer UM. Altered resting state brain networks in Parkinson's disease. PloS one. 2013 Oct 28; 8(10):e77336. https://doi.org/10.1371/journal.pone.0077336 PMID: 24204812

4. Long D, Wang J, Xuan M, Gu Q, Xu X, Kong D, et al. Automatic classification of early Parkinson's disease with multi-modal MR imaging. PloS one. 2012 Nov 9; 7(11):e47714. https://doi.org/10.1371/journal.pone.0047714 PMID: 23152757

5. Skidmore FM, Yang M, Baxter L, Von Deneen KM, Collingwood J, He G, et al. Reliability analysis of the resting state can sensitively and specifically identify the presence of Parkinson disease. Neuroimage. 2013 Jul 15; 75:249–61. https://doi.org/10.1016/j.neuroimage.2011.06.056 PMID: 21924367

6. Baudrexel S, Witte T, Seifried C, von Wegner F, Beissner F, Klein JC, et al. Resting state fMRI reveals increased subthalamic nucleus—motor cortex connectivity in Parkinson's disease. Neuroimage. 2011 Apr 15; 55(4):1728–38. https://doi.org/10.1016/j.neuroimage.2011.01.017 PMID: 21255661

7. Wu T, Long X, Wang L, Hallett M, Zang Y, Li K, et al. Functional connectivity of cortical motor areas in the resting state in Parkinson's disease. Human brain mapping. 2011 Sep 1; 32(9):1443–57. https://doi.org/10.1002/hbm.21118 PMID: 20740649

8. Kwak Y, Peltier S, Bohnen N, Müller M, Dayalu P, Seidler RD. Altered resting state cortico-striatal connectivity in mild to moderate stage Parkinson's disease. Frontiers in systems neuroscience. 2010 Sep 15; 4:143. https://doi.org/10.3389/fnsys.2010.00143 PMID: 21206528

9. Kwak Y, Peltier S, Bohnen N, Müller M, Dayalu P, Seidler RD. L-DOPA changes spontaneous low-frequency BOLD signal oscillations in Parkinson's disease: a resting state fMRI study. Frontiers in systems neuroscience. 2012 Jul 4; 6:52. https://doi.org/10.3389/fnsys.2012.00052 PMID: 22783172

10. Helmich RC, Derikx LC, Bakker M, Scheeringa R, Bloem BR, Toni I. Spatial remapping of cortico-striatal connectivity in Parkinson's disease. Cerebral cortex. 2010 May 1; 20(5):1175–86. https://doi.org/10.1093/cercor/bhp178 PMID: 19710357

11. Helmich RC, Janssen MJ, Oyen WJ, Bloem BR, Toni I. Pallidal dysfunction drives a cerebellothalamic circuit into Parkinson tremor. Annals of neurology. 2011 Feb 1; 69(2):269–81. https://doi.org/10.1002/ana.22361 PMID: 21387372

12. Luo C, Song W, Chen Q, Zheng Z, Chen K, Cao B, et al. Reduced functional connectivity in early-stage drug-naive Parkinson's disease: a resting-state fMRI study. Neurobiology of aging. 2014 Feb 28; 35 (2):431–41. https://doi.org/10.1016/j.neurobiolaging.2013.08.018 PMID: 24074808

13. Yu R, Liu B, Wang L, Chen J, Liu X. Enhanced functional connectivity between putamen and supplementary motor area in Parkinson's disease patients. PLoS One. 2013 Mar 26; 8(3):e59717. https://doi.org/10.1371/journal.pone.0059717 PMID: 23555758

14. Hacker C, Perlmutter J, Criswell S, Ances B, Snyder A. Resting state functional connectivity of the striatum in Parkinson's disease. Brain. 2012 Nov 28:aws281.

15. Kurani AS, Seidler RD, Burciu RG, Comella CL, Corcos DM, Okun MS, et al. Subthalamic nucleus—sensorimotor cortex functional connectivity in de novo and moderate Parkinson's disease. Neurobiology of aging. 2015 Jan 31; 36(1):462–9. https://doi.org/10.1016/j.neurobiolaging.2014.07.004 PMID: 25095723

16. Baggio HC, Segura B, Sala-Llonch R, Marti MJ, Valldeoriola F, Compta Y, et al. Cognitive impairment and resting-state network connectivity in Parkinson's disease. Human brain mapping. 2015 Jan 1; 36 (1):199–212. https://doi.org/10.1002/hbm.22622 PMID: 25164875

17. Esposito F, Tessitore A, Giordano A, De Micco R, Paccone A, Conforti R, et al. Rhythm-specific modulation of the sensorimotor network in drug-naive patients with Parkinson's disease by levodopa. Brain. 2013 Mar 1; 136(3):710–25.

18. Szewczyk-Krolikowski K, Menke RA, Rolinski M, Duff E, Salimi-Khorshidi G, Filippini N, et al. Functional connectivity in the basal ganglia network differentiates PD patients from controls. Neurology. 2014 Jul 15; 83(3):208–14. https://doi.org/10.1212/WNL.0000000000000592 PMID: 24920856

19. Tessitore A, Amboni M, Esposito F, Russo A, Picillo M, Marcuccio L, et al. Resting-state brain connectivity in patients with Parkinson's disease and freezing of gait. Parkinsonism & related disorders. 2012 Jul 31; 18(6):781–7.

20. Sharman M, Valabregue R, Perlbarg V, Marrakchi-Kacem L, Vidailhet M, Benali H, et al. Parkinson's disease patients show reduced cortical-subcortical sensorimotor connectivity. Movement Disorders. 2013 Apr 1; 28(4):447–54. https://doi.org/10.1002/mds.25255 PMID: 23144002

21. Wu T, Wang J, Wang C, Hallett M, Zang Y, Wu X, et al. Basal ganglia circuits changes in Parkinson's disease patients. Neuroscience letters. 2012 Aug 22; 524(1):55–9. https://doi.org/10.1016/j.neulet.2012.07.012 PMID: 22813979

22. Liu H, Edmiston EK, Fan G, Xu K, Zhao B, Shang X, et al. Altered resting-state functional connectivity of the dentate nucleus in Parkinson's disease. Psychiatry Research: Neuroimaging. 2013 Jan 30; 211 (1):64–71. https://doi.org/10.1016/j.pscychresns.2012.10.007 PMID: 23352277

23. Chen Y, Yang W, Long J, Zhang Y, Feng J, Li Y, et al. Discriminative Analysis of Parkinson's Disease Based on Whole-Brain Functional Connectivity. PloS one. 2015 Apr 17; 10(4):e0124153. https://doi.org/10.1371/journal.pone.0124153 PMID: 25885059

24. Wen X, Wu X, Liu J, Li K, Yao L. Abnormal baseline brain activity in non-depressed Parkinson's disease and depressed Parkinson's disease: a resting-state functional magnetic resonance imaging study. PloS one. 2013 May 22; 8(5):e63691. https://doi.org/10.1371/journal.pone.0063691 PMID: 23717467

25. Zhang D, Liu X, Chen J, Liu B, Wang J. Widespread increase of functional connectivity in Parkinson's disease with tremor: a resting-state FMRI study. Frontiers in aging neuroscience. 2015; 7.

26. Tahmasian M, Bettray LM, van Eimeren T, Drzezga A, Timmermann L, Eickhoff CR, et al. A systematic review on the applications of resting-state fMRI in Parkinson's disease: Does dopamine replacement therapy play a role?. Cortex. 2015 Dec 31; 73:80–105. https://doi.org/10.1016/j.cortex.2015.08.005 PMID: 26386442

27. Power JD, Schlaggar BL, Petersen SE. Recent progress and outstanding issues in motion correction in resting state fMRI. Neuroimage. 2015 Jan 15; 105:536–51. https://doi.org/10.1016/j.neuroimage.2014.10.044 PMID: 25462692

28. Friston KJ, Buchel C. Functional connectivity. Human Brain Function. 2003; 2:999–1018.

29. Shirer WR, Ryali S, Rykhlevskaia E, Menon V, Greicius MD. Decoding subject-driven cognitive states with whole-brain connectivity patterns. Cerebral cortex. 2012 Jan 1; 22(1):158–65. https://doi.org/10.1093/cercor/bhr099 PMID: 21616982

30. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage. 2002 Jan 31; 15(1):273–89. https://doi.org/10.1006/nimg.2001.0978 PMID: 11771995

31. Craddock RC, James GA, Holtzheimer PE, Hu XP, Mayberg HS. A whole brain fMRI atlas generated via spatially constrained spectral clustering. Human brain mapping. 2012 Aug 1; 33(8):1914–28. https://doi.org/10.1002/hbm.21333 PMID: 21769991

32. Shen X, Tokoglu F, Papademetris X, Constable RT. Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. Neuroimage. 2013 Nov 15; 82:403–15. https://doi.org/10.1016/j.neuroimage.2013.05.081 PMID: 23747961

33. Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. Journal of cognitive neuroscience. 2007 Sep; 19(9):1498–507. https://doi.org/10.1162/jocn.2007.19.9.1498 PMID: 17714011

34. Power JD, Cohen AL, Nelson SM, Wig GS, Barnes KA, Church JA, et al. Functional network organization of the human brain. Neuron. 2011 Nov 17; 72(4):665–78. https://doi.org/10.1016/j.neuron.2011.09.006 PMID: 22099467

35. Gordon EM, Laumann TO, Adeyemo B, Huckins JF, Kelley WM, Petersen SE. Generation and evaluation of a cortical area parcellation from resting-state correlations. Cerebral cortex. 2014 Oct 14:bhu239.

36. Talairach J, Tournoux P. Co-planar stereotaxic atlas of the human brain. 3-Dimensional proportional system: an approach to cerebral imaging. 1988

37. Murphy K, Birn RM, Handwerker DA, Jones TB, Bandettini PA. The impact of global signal regression on resting state correlations: are anti-correlated networks introduced?. Neuroimage. 2009 Feb 1; 44(3):893–905. https://doi.org/10.1016/j.neuroimage.2008.09.036 PMID: 18976716

38. Hayasaka S. Functional connectivity networks with and without global signal correction. Frontiers in human neuroscience 7. 2013.

39. Liu D, Dong Z, Zuo X, Wang J, Zang Y. Eyes-open/eyes-closed dataset sharing for reproducibility evaluation of resting state fMRI data analysis methods. Neuroinformatics. 2013 Oct 1; 11(4):469–76. https://doi.org/10.1007/s12021-013-9187-0 PMID: 23836389

40. Griffanti L, Salimi-Khorshidi G, Beckmann CF, Auerbach EJ, Douaud G, Sexton CE, et al. ICA-based artefact removal and accelerated fMRI acquisition for improved resting state network imaging. Neuroimage. 2014 Jul 15; 95:232–47. https://doi.org/10.1016/j.neuroimage.2014.03.034 PMID: 24657355

41. Salimi-Khorshidi G, Douaud G, Beckmann CF, Glasser MF, Griffanti L, Smith SM. Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. Neuroimage. 2014 Apr 15; 90:449–68. https://doi.org/10.1016/j.neuroimage.2013.11.046 PMID: 24389422

42. Xia M, Wang J, He Y. BrainNet Viewer: a network visualization tool for human brain connectomics. PloS one. 2013 Jul 4; 8(7):e68910. https://doi.org/10.1371/journal.pone.0068910 PMID: 23861951

43. Brunet JP, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. Proceedings of the national academy of sciences. 2004 Mar 23; 101(12):4164–9.

44. Orban P, Dansereau C, Desbois L, Mongeau-Pérusse V, Giguère CÉ, Nguyen H, et al. Multisite generalizability of schizophrenia diagnosis classification based on functional brain connectivity. Schizophrenia research. 2017 Jun 7.

45. Wu T, Ma Y, Zheng Z, Peng S, Wu X, Eidelberg D, et al. Parkinson's Disease—Related Spatial Covariance Pattern Identified with Resting-State Functional MRI. Journal of Cerebral Blood Flow & Metabolism. 2015 Nov 1; 35(11):1764–70.