

# SCIENTIFIC REPORTS



OPEN

## Imaging genetics approach to Parkinson's disease and its correlation with clinical score

Mansu Kim<sup>1,2</sup>, Jonghoon Kim<sup>1,2</sup>, Seung-Hak Lee<sup>1,2</sup> & Hyunjin Park<sup>2,3</sup>

Received: 07 November 2016

Accepted: 24 March 2017

Published: 21 April 2017

Parkinson's disease (PD) is a progressive neurodegenerative disorder associated with both underlying genetic factors and neuroimaging findings. Existing neuroimaging studies related to the genome in PD have mostly focused on certain candidate genes. The aim of our study was to construct a linear regression model using both genetic and neuroimaging features to better predict clinical scores compared to conventional approaches. We obtained neuroimaging and DNA genotyping data from a research database. Connectivity analysis was applied to identify neuroimaging features that could differentiate between healthy control (HC) and PD groups. A joint analysis of genetic and imaging information known as imaging genetics was applied to investigate genetic variants. We then compared the utility of combining different genetic variants and neuroimaging features for predicting the Movement Disorder Society-sponsored unified Parkinson's disease rating scale (MDS-UPDRS) in a regression framework. The associative cortex, motor cortex, thalamus, and pallidum showed significantly different connectivity between the HC and PD groups. Imaging genetics analysis identified PARK2, PARK7, HtrA2, GIGYRF2, and SNCA as genetic variants that are significantly associated with imaging phenotypes. A linear regression model combining genetic and neuroimaging features predicted the MDS-UPDRS with lower error and higher correlation with the actual MDS-UPDRS compared to other models using only genetic or neuroimaging information alone.

Parkinson's disease (PD) is a progressive neurodegenerative disorder characterized by bradykinesia, resting tremors, rigidity, and difficulty with voluntary movement<sup>1</sup>. Typically, PD is diagnosed using criteria such as the United Kingdom's Parkinson's Disease Society Brain Bank<sup>2</sup>. The pathological hallmark of PD is the loss of dopaminergic neurons in the substantia nigra (SN) and those that project into the striatum. This neuronal loss leads to dopamine imbalance, which may result in inhibition of basal ganglia output and dysfunction within the cortico-basal ganglia-thalamocortical (CBGT) circuit<sup>3,4</sup>. Environmental toxins and genetic factors have been linked to the development of PD<sup>5</sup>. More than 16 PD-related genetic mutations have been recognized<sup>6</sup>.

Various neuroimaging approaches have been used to investigate how PD affects the brain<sup>4,7–10</sup>. These imaging approaches include positron emission tomography (PET), single-photon emission computed tomography (SPECT), and magnetic resonance imaging (MRI). Functional dopaminergic imaging such as dopamine transporter SPECT has been widely used to assess nigrostriatal degeneration<sup>11</sup>. White matter integrity could be assessed using diffusion MRI and tractography, which allows investigation of neuronal degeneration<sup>12</sup>. These various neuroimaging approaches allow us to better understand the neurobiological mechanism of PD<sup>13</sup>.

Existing neuroimaging studies related to the genome of PD have focused on certain candidate genes<sup>14,15</sup>. Recently, genome-wide association study (GWAS) has been used to identify genetic polymorphisms that influence disease phenotypes<sup>16</sup>. Common genetic variants that are associated with many diseases such as Alzheimer's disease, PD, and diabetes have been successfully identified using GWAS<sup>16,17</sup>. Genes associated with PD have been identified using GWAS<sup>6,18</sup>. GWAS is frequently limited by the need for a large number of data samples to achieve statistical significance. Sample size requirements can be reduced using quantitative trait analysis, a form of genetic analysis that uses intermediate phenotypes of the disorder being studied<sup>19</sup>. Imaging genetics is a type of quantitative trait analysis that uses imaging measures (e.g., functional connectivity and brain activation) that are heritable and correlated with the disease in question.

<sup>1</sup>Department of Electronic, Electrical and Computer Engineering, Sungkyunkwan University, Korea. <sup>2</sup>Center for Neuroscience Imaging Research, Institute for Basic Science, Korea. <sup>3</sup>School of Electronic and Electrical Engineering, Sungkyunkwan University, Korea. Correspondence and requests for materials should be addressed to H.P. (email: hyunjinp@skku.edu)

Region	HC	PD	Corrected p-value
Associative cortex	0.017 ± 0.004	0.019 ± 0.004	0.047
Motor cortex	0.012 ± 0.003	0.015 ± 0.003	0.022
Thalamus	0.007 ± 0.002	0.008 ± 0.002	<0.001
Pallidum	0.006 ± 0.001	0.005 ± 0.001	0.007

**Table 1.** DC values of regions with significantly different structural connectivity in PD patients versus HC subjects. DC values are reported as the mean ± standard deviation (SD). Corrected p-values are reported in the right-hand column.

Imaging genetics is an emerging tool that is capable of identifying the genetic risk of a disease based on both clinical factors and imaging techniques<sup>20</sup>. Imaging techniques provide complementary high dimensional information regarding the disease, and thus, GWAS could be improved by such additional information. Existing imaging genetic studies have been applied to Alzheimer's disease and schizophrenia<sup>17,19,21–23</sup>. However, the application of imaging genetics for PD has been scarce. Prior imaging genetics studies have focused on specific gene candidates and have not used imaging measures for quantitative trait analysis<sup>14,15</sup>. One study investigating the cause of familial PD using functional PET imaging and genetic data reported that a mutation in the LRRK2 gene led to familial PD<sup>14</sup>. Another study investigated the early progression of PD in patients with mutant Parkin and PINK genes using functional MRI (fMRI)<sup>15</sup>. They showed that Parkin mutation carriers had abnormalities in the supplementary motor area and premotor cortex during a simple sequence task and that PINK mutation carriers were affected in the same regions during a finger sequence task.

In our study, we applied imaging genetics combining genetic polymorphisms and brain connectivity measurements as intermediate phenotypes to identify clinically relevant genetic variants in PD. Another aim of our study was to determine if the use of genetic variants along with their associated neuroimaging features could better predict clinical scores of PD progression compared to using conventional approaches with GWAS or neuroimaging alone. We obtained diffusion MRI and DNA genotyping data from the Parkinson's disease Progression Marker Initiative (PPMI) database<sup>24</sup>. Connectivity analysis was used to identify neuroimaging features within the CBCT circuit that could differentiate between healthy control (HC) and PD groups. These neuroimaging features and associated clinical scores were used as intermediate phenotypes in imaging genetics analysis to identify clinically relevant genetic variants. We then analyzed the utility of combining different genetic variants and neuroimaging features for predicting the Movement Disorder Society-sponsored unified Parkinson's disease rating scale (MDS-UPDRS), a well-established measure reflecting motor symptoms and clinimetric properties of PD, using a regression framework<sup>25</sup>.

## Results

**Subject selection.** First, quality control of genotype data was performed based on the Enhancing Neuro Imaging Genetics through Meta Analysis (ENIGMA) protocol. In total, single-nucleotide polymorphisms (SNPs) with a minor allele frequency <1% (n = 53,482), Hardy-Weinberg equilibrium <10<sup>-6</sup> (n = 683), and a genotype missing rate >5% (n = 14,378) were excluded from the study. Following quality control, 136,350 SNPs remained for further analysis. Second, the population stratification of the genetic data was adjusted using multidimensional scaling. After population stratification, only unrelated Caucasian subjects (n = 418) remained (Supplementary Figure S1). Of these unrelated Caucasian subjects, 146 (HC = 40, PD = 106) had both neuroimaging and genetic data available in the database. Forty of these PD cases were then randomly chosen to achieve equal numbers of samples in each group.

**Connectivity differences.** Differences between the HC and PD groups were quantified in terms of degree centrality (DC) values using t-tests with Bonferroni correction. Four regions, namely the associative cortex, motor cortex, thalamus, and pallidum, had significantly different DC values between the HC and PD groups (Table 1). The associative cortex, motor cortex, and thalamus showed increased DC values in PD patients compared with HC subjects, while the pallidum showed reduced DC values in PD patients compared with HC subjects.

**Genetic variants in imaging genetics analysis.** The results of imaging genetics analysis using the DC values of the associative cortex, motor cortex, thalamus, and pallidum and MDS-UPDRS are shown in Table 2. Three SNPs were associated with the DC values of these four regions, and four SNPs were associated with the MDS-UPDRS. PARK2 (which contained three SNPs) was associated with the DC value of the associative cortex and MDS-UPDRS. PARK7 was associated with the DC value of the motor cortex, and SNCA was associated with the DC value of the thalamus. HtrA2 and GIGYF2 were associated with the MDS-UPDRS.

**Linear model analysis.** We compared three multiple linear regression models to determine which model best explained the MDS-UPDRS. Stepwise regression was applied to construct three linear regression models, and the process led to different regression models depending on the initial model. *Model C* (the one with both genetic variants and neuroimaging features) (3) showed the best performance (adjusted R-squared (adj-R<sup>2</sup>) = 0.714, p < 0.001) compared to *model A* (the one with neuroimaging features) (1) and *model B* (the one with genetic variants) (2), which yielded adj-R<sup>2</sup> values of 0.508 (p < 0.001) and 0.442 (p < 0.001), respectively (Tables 3, 4 and 5). Details regarding the three equations are found in the Methods section. Specifically, the DC value of the pallidum singly explained the greatest (9.70%) variance (F = 21.52, p < 0.001) in *model C* (3). *Model C* (3) explained 72.51% of the variance in the MDS-UPDRS, while *model A* (1) and *model B* (2) explained 53.25% and 42.74% of

Intermediate phenotype	Gene	SNP	Base-pair position	Minor allele	Empirical p-value
Associative cortex	PARK2	rs6901583	162543210	C	0.015
Motor cortex	PARK7	rs225103	8042301	T	0.043
Thalamus	SNCA	rs1473533	90735702	T	0.049
MDS-UPDRS	PARK2	rs9346876	162032871	C	0.034
	PARK2	rs9364608	162034799	G	0.034
	HtrA2	rs363611	74759708	T	0.044
	GIGYF2	rs7421653	233711471	G	0.048

**Table 2.** Genetic variants identified by imaging genetics analysis using neuroimaging features and clinical score as intermediate phenotypes.

	Sum of squares	F-statistic	P-value	Pct. Exp. (%)
Associative cortex	682.07	8.65	0.004	5.78
Thalamus	1268.30	16.09	<0.001	10.75
Pallidum	980.81	12.44	<0.001	8.31
Family history	1023.50	12.98	<0.001	8.67
GDS	408.72	5.18	0.025	3.46
Age	190.98	2.42	0.124	1.62
Associative × Family history	710.76	9.01	0.004	6.02
Thalamus × GDS	609.60	7.73	0.007	5.16
Family history × Age	410.49	5.20	0.025	3.48
<b>Total</b>				53.25

**Table 3.** Detailed results of model A equation (1). Pct. Exp.: percentage of explained variance.

	Sum of squares	F-statistic	P-value	Pct. Exp. (%)
rs6901583 (PARK2)	83.13	0.83	0.364	0.67
rs1473533 (SNCA)	136.13	1.36	0.247	1.10
rs9346876 (PARK2)	348.26	3.49	0.066	2.81
rs363611 (HtrA2)	703.45	7.04	0.009	5.68
GDS	303.08	3.03	0.085	2.45
Sex	523.84	5.24	0.025	4.23
rs1473533 × rs363611	1303.70	13.05	<0.001	10.52
rs9346876 × rs363611	1893.50	18.95	<0.001	15.28
<b>Total</b>				42.74

**Table 4.** Detailed results of model B equation (2). Pct. Exp.: percentage of explained variance.

the variance in the MDS-UPDRS (Tables 3,4 and 5). Thus, combining genetic variants with neuroimaging features resulted in an improved adj- $R^2$  value, indicating that they provide complementary information. Each model included a distinct set of independent variables (IVs), and thus, the number of included IVs varied from model to model. A given variable might appear in two different models, but its contribution in terms of percentage of explained variance could be different as there was a distinct set of IVs in each model. For example, the DC of the thalamus explained 10.75% of the variance in conjunction with 5 other IVs in *model A*, while the DC of the same region explained 6.87% in conjunction with 9 other IVs in *model C*.

**Prediction of MDS-UPDRS.** The model combining genetic and neuroimaging features, *model C* (3), predicted the MDS-UPDRS with root mean squared error (RMSE) of 7.82 and showed a high correlation with the actual MDS-UPDRS ( $r = 0.788$ ,  $p < 0.001$ ) compared with the models using genetic and neuroimaging features alone (*model A* (1) and *model B* (2)). The first model with only neuroimaging features, *model A* (1), showed a high RMSE of 10.24 and a correlation of 0.616 ( $p < 0.001$ ). The second model with only genetic variants, *model B* (2), yielded a RMSE of 11.71 and a correlation of 0.475 ( $p < 0.001$ ). Plots of the actual and predicted MDS-UPDRS scores for these three models are given in Fig. 1.

## Discussion

Our main finding showed that the linear prediction model combining neuroimaging and genetics could explain 72.51% of the variance in the MDS-UPDRS, and the model resulted in a high correlation ( $r = 0.788$ ) between

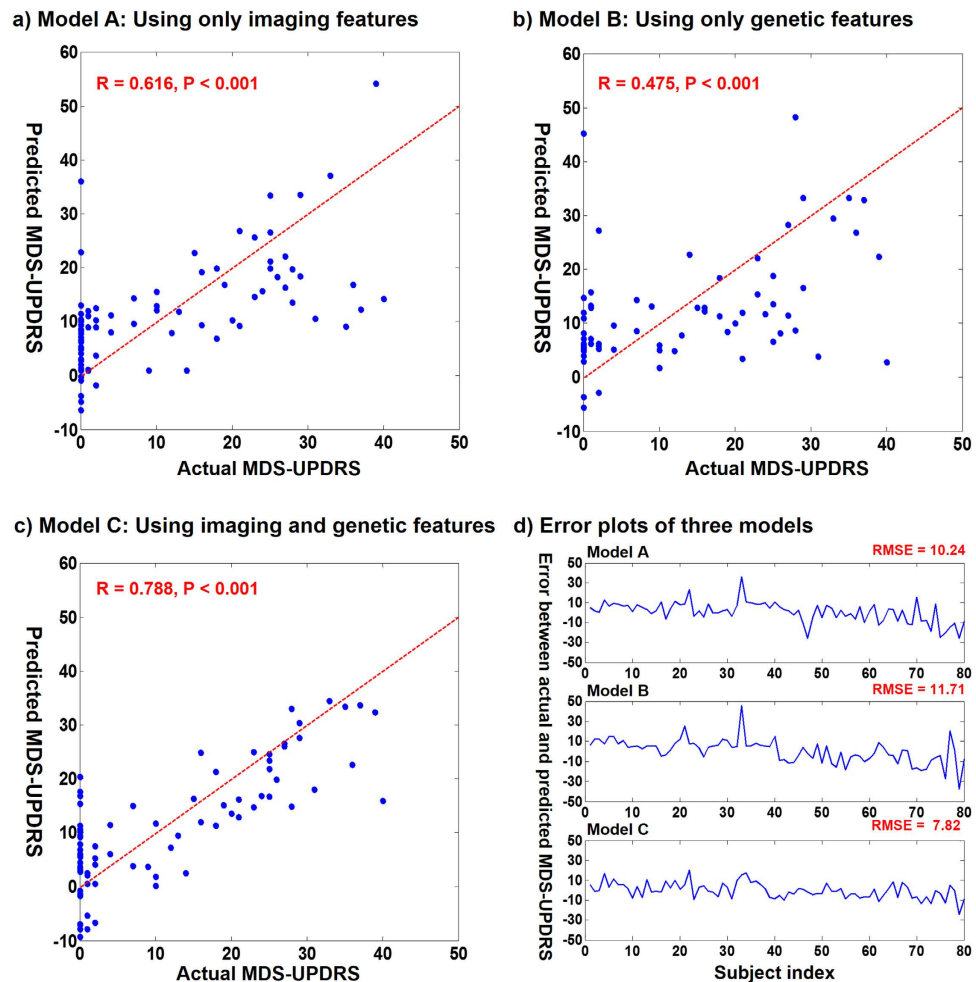
	Sum of squares	F-statistic	P-value	Pct. Exp. (%)
Motor cortex	472.94	10.30	0.002	4.64
Thalamus	699.97	15.24	<0.001	6.87
Pallidum	988.10	21.52	<0.001	9.70
rs6901583 (PARK2)	24.39	0.53	0.469	0.24
rs9346876 (PARK2)	649.43	14.14	<0.001	6.37
rs7421653 (GIGYF2)	135.18	2.94	0.091	1.33
rs363611 (HtrA2)	277.45	6.04	0.017	2.72
GDS	200.06	4.36	0.041	1.96
Sex	436.84	9.51	0.003	4.29
Age	228.45	4.98	0.029	2.24
Thalamus × rs7421653	430.90	9.38	0.003	4.23
Pallidum × rs6901583	246.29	5.36	0.024	2.42
Pallidum × Age	217.02	4.73	0.034	2.13
rs6901583 × rs363611	448.84	9.78	0.003	4.41
rs9346876 × rs7421653	132.27	2.88	0.095	1.30
rs9346876 × rs363611	927.71	20.20	<0.001	9.11
rs6901583 × GDS	496.61	10.82	0.002	4.87
rs363611 × GDS	375.28	8.17	0.006	3.68
<b>Total</b>				<b>72.51</b>

**Table 5. Detailed results of model C equation (3).** Pct. Exp.: percentage of explained variance.

the predicted and actual MDS-UPDRS. Our connectivity analysis revealed that the associative cortex, motor cortex, thalamus, and pallidum had significantly different DC values between the HC and PD groups. Imaging genetics analysis identified PARK2, PARK7, HtrA2, GIGYF2, and SNCA as genetic variants that are significantly associated with intermediate phenotypes. These findings illustrate that the identified PD-related genes might yield structural connectivity changes that could be recognized using diffusion MRI within the CBGT circuit. The linear regression model combining genetic and neuroimaging features predicted the MDS-UPDRS with a lower RMSE and higher correlation with the actual MDS-UPDRS compared to the other models using only genetic or neuroimaging information alone.

Many recent neuroimaging studies have investigated PD using brain connectivity<sup>4,8,10,26,27</sup>. Connectivity analysis assumes that the brain is a complex interconnected network, and it is suitable for exploring large-scale circuits that include the cortical-thalamic and cerebro-cerebellar circuits<sup>28</sup>. Connectivity could be assessed using centrality measures including betweenness, eigenvector, degree, and page-rank centrality to assess network properties in the brain<sup>29</sup>. Each centrality characterizes the importance of a given node in terms of global or local contribution to the network. There is no single centrality measure that works well in all network assessments, and the optimal centrality measure is different depending on the imaging modality and the disease of concern<sup>30</sup>. DC has a straightforward neurobiological interpretation in structural connectivity networks, and thus, we utilized DC in this study. Our connectivity analysis revealed that the associative cortex, motor cortex, thalamus, and pallidum had significantly different DC values between the HC and PD groups. Some previous studies also showed that these regions were important in PD patients based on fMRI, diffusion MRI, and SPECT<sup>8,27,31–33</sup>.

Conventional genetic analyses have identified genetic variants linked with PD<sup>5,18</sup>. Imaging genetics combines imaging information with genetic information, allowing for identification of additional genetic variants<sup>22</sup>. Thus, the analysis of PD with imaging genetics might identify additional information to better investigate the neurobiological mechanisms of PD. In this study, an effect of PARK2 (Parkin) was observed on the DC value of the associative cortex and clinical scores. PARK2 plays an important role in presynaptic and postsynaptic neuronal processes<sup>34</sup>. Mutations in the Parkin protein are responsible for disrupting E3 ubiquitin ligase activity in PD<sup>35</sup>. The associative cortex plays a secondary role in processing motor information in the classic model of the CBGT motor circuit<sup>31</sup>. Previous work showed that a mutation of PARK2 was associated with activity in the left lateral orbitofrontal cortex, which is part of the associative cortex<sup>36</sup>. Another study reported that the DC value of the dorsolateral prefrontal sub-region of the associative cortex was increased in PD patients compared to HC using resting state-fMRI<sup>8</sup>. These findings suggest that PARK2 may be linked with connectivity in PD patients and in turn may affect the MDS-UPDRS. An effect of PARK7 (DJ-1) was observed on DC values in the motor cortex. PARK7 is involved in transcriptional regulation and response to oxidative stress<sup>37</sup>. The motor cortex is a key region involved in motor processing, serving as a major input into the basal ganglia<sup>1,13</sup>. Our findings suggest that PARK7 may affect neurons in the motor cortex and in turn may affect PD symptoms. Previous reports showed that characteristic PD findings, such as dopaminergic neuronal death in the SN and motor abnormalities, are affected by mutations in PARK7<sup>37</sup>. One resting state-fMRI study showed that connectivity between the putamen and motor cortex increased in PD patients<sup>27</sup>. Another study reported that the DC value of the parietal cortex, a sub-region of the motor cortex, was increased in PD patients compared with HC subjects<sup>8</sup>. Moreover, one diffusion MRI study reported that track density in the primary somatosensory cortex was increased in PD patients compared with HC subjects<sup>32</sup>. Thus, the physiologic association of PARK7 with neurons in the motor cortex suggested by our results could be feasible. The interaction term between rs9346876 (PARK2) and rs363611 (HtrA2)



**Figure 1.** Plots of actual and predicted MDS-UPDRS for three prediction models. Sub-figures (a,b and c) show actual and predicted MDS-UPDRS from *Models A* (1), *B* (2), and *C* (3), respectively. The dashed line indicates the identity line (i.e., actual score = predicted score). (d) shows error plots of the three models.

led to 15.28% and 9.11% of explained variance for *models B* and *C*, respectively, in the linear regression analysis. This finding is consistent with previous studies reporting that Htra2 specifically binds to and directly cleaves the E3 ubiquitin ligase PARK2<sup>38,39</sup>.

We performed additional analysis with 14 right-sided and 14 left-sided onset cases in the PD group using contralateral ROIs, and detailed results are shown in the supplemental material. We observed better results for the reduced 28 patient group compared to using the previous 40 patients in terms of adj-R<sup>2</sup> values. *Model C*'s adj-R<sup>2</sup> improved from 0.714 to 0.768, *model A*'s adj-R<sup>2</sup> improved from 0.508 to 0.650, and *model B*'s adj-R<sup>2</sup> improved from 0.422 to 0.485. Similar to the main results with 40 patients, *model C* (adj-R<sup>2</sup> = 0.768) outperformed models *A* (adj-R<sup>2</sup> = 0.650) and *B* (adj-R<sup>2</sup> = 0.485) implying that combining genetic variants with neuroimaging features indeed was the best model. Similar to the main results, for the reduced 28 patients, *model C* showed the best prediction performance with low RMSE (RMSE = 3.35) and a high correlation with the actual MDS-UPDRS ( $r = 0.847, p < 0.001$ ).

Our study was limited by the small number of available samples. This issue was partly mitigated by analyzing SNPs and genes previously linked with PD<sup>5</sup>. Our work focused on known genetic variants, thereby establishing a foundation for future research. The neuroimaging data in our study came from nine different clinical sites, but all nine sites adopted the same image acquisition protocol using the same MRI scanner model (Siemens 3T scanner) to reduce confounding effects. In addition, we added multi-center information as a nuisance co-variate to account for multi-center confounding effects. We restricted the connectivity analysis to eight known regions within the CBGT circuit, which did not cover the entire brain. We used a single imaging modality (i.e., diffusion MRI) to compute intermediate phenotypes of PD. A multi-modal imaging study, potentially including PET and SPECT, could provide complementary information and thus has the potential to provide further insight into the pathophysiology of PD. Finally, we retrieved data from a research database that lacked follow-up longitudinal data. Future work should consider longitudinal data to test the validity of our models.

## Methods

**Subjects and imaging acquisition.** This study was a retrospective analysis of anonymized data, and institutional review board (IRB) approval was obtained at Sungkyunwan University. Our study was performed in full accordance with the local IRB guidelines. Informed consent was obtained from all subjects. The baseline neuroimaging (T1-, T2-, and diffusion MR) and DNA genotyping data were obtained from the PPMI research database<sup>24</sup>. In detail, diffusion MRI scans were acquired with the following acquisition parameters: TR = 900 ms, TE = 88 ms, b = 1000 s/mm<sup>2</sup>, 64 diffusion gradient directions with one b<sub>0</sub> image, image matrix = 116 × 116 × 72, and voxel resolution = 1.98 × 1.98 × 2 mm<sup>3</sup>. T1-weighted MRI scans were acquired using the following parameters: TR = 2300 ms, TE = 2.98 ms, TI = 900 ms, image matrix = 240 × 256 × 176, and voxel resolution = 1 × 1 × 1 mm<sup>3</sup>. T2-weighted MRI scans were acquired using the following parameters: TR = 3000 ms, TE = 101 ms, image matrix = 228 × 256 × 48, and voxel resolution = 0.9375 × 0.9375 × 3 mm<sup>3</sup>. The PPMI database included 423 PD and 196 HC cases as of July 2015. PD cases were classified using the criteria established by the PPMI consortium<sup>24</sup>. Subjects who underwent both MRI imaging and genetic testing were considered for this study. We used off-medication subjects so that the effects of levodopa therapy were excluded. Additional patient data including MDS-UPDRS, Hohn and Yahn, Geriatric Depression Scale (GDS), Montreal Cognitive Assessment (MoCA), handedness and family history were collected for each case at the baseline visit (Supplementary Table S1). Population stratification of genetic data was adjusted using multidimensional scaling. We used multidimensional scaling analysis based on the ENIGMA protocol to reduce the population stratification and to ensure genetic homogeneity for all subjects<sup>40</sup>. All quality controlled SNPs were compared to the HapMap3 reference population using Plink v1.09 software<sup>41</sup>. Identity by state distance was calculated for all autosomal SNPs so that only unrelated Caucasian subjects remained. Of these unrelated Caucasian subjects, we chose 40 HC and 40 PD cases as shown in the Results section. The age, sex ratio, handedness, and cognitive assessments of the HC and PD groups were matched (Supplementary Table S1). Among 40 PD patients, 14 were right-sided, 14 were left-sided, and 12 were bilateral onset cases.

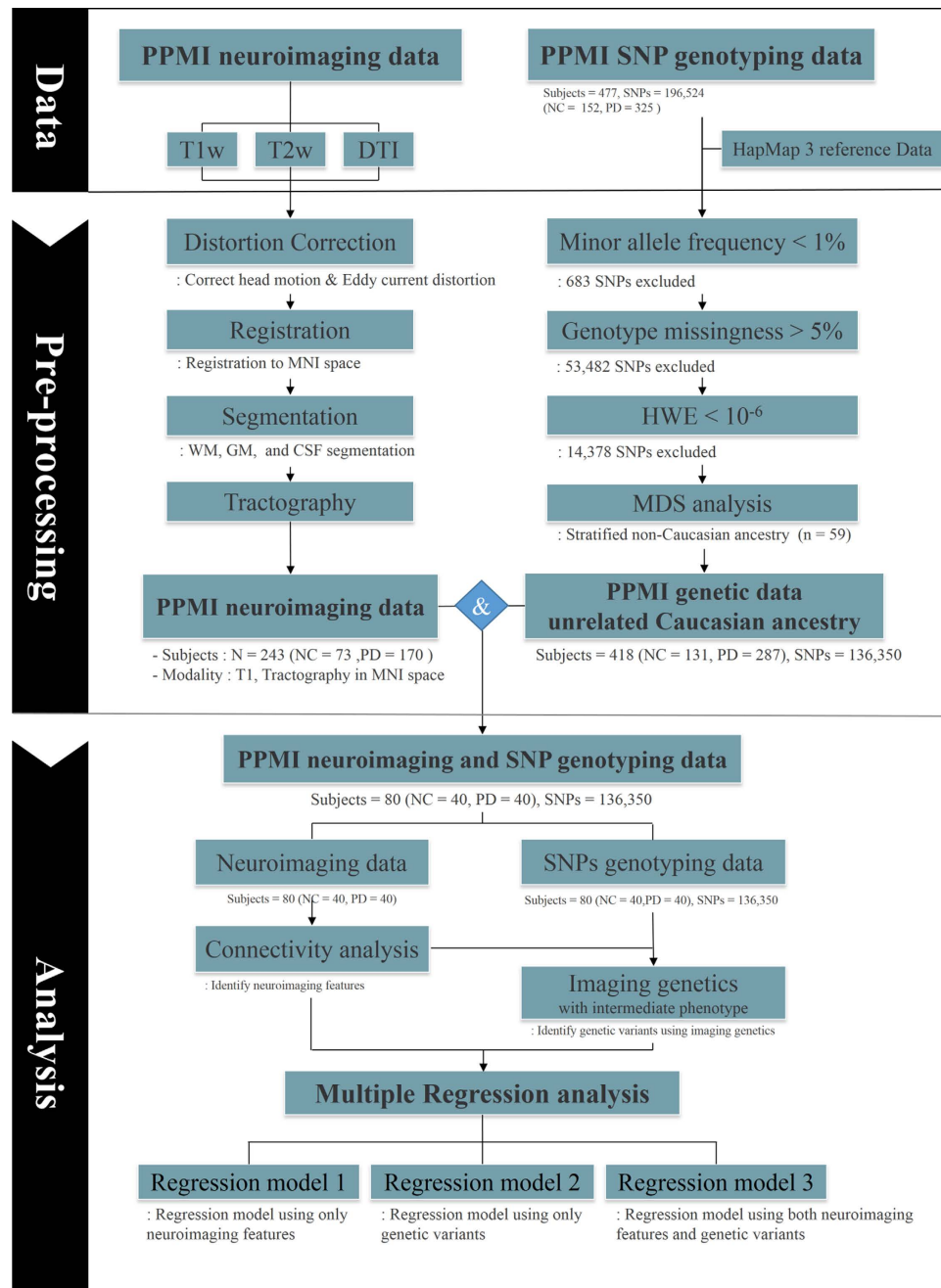
**Tractography.** Diffusion MRI data were processed with tractography, an algorithm that can extract *in vivo* neuronal fiber information, as described in Fig. 2. In brief, distortions due to eddy currents and head motion were first corrected using FSL. Second, T1-, T2-, and diffusion-weighted MRI scans were aligned to the Montreal Neurological Institute space with non-linear registration using FSL<sup>42</sup>. Then, white matter, gray matter, and cerebrospinal fluid were segmented from the registered T1-weighted image using Freesurfer<sup>43</sup>. The segmented white matter was used to guide the tractography algorithm. Fiber information was computed using the deterministic tractography algorithm implemented in the Diffusion Toolkit<sup>44</sup>. For connectivity analysis (performed with MATLAB), structural connectivity matrices were constructed using a deterministic tractography algorithm.

**Region of interest specification.** Connectivity analysis requires nodes and edges in a graph structure. Eight ROIs associated with PD were adopted from the existing literature and used as connectivity nodes<sup>4</sup>. The eight ROIs were the caudate, putamen, pallidum, thalamus, sensorimotor cortex, associative cortex, limbic cortex (specified by the Desikan-Killiany atlas) and SN (specified by transferring ROI information from a pre-defined atlas via image co-registration) (Fig. 3)<sup>45,46</sup>. These regions are members of the CBGT circuit and are known to be associated with motor, executive, and movement functions<sup>1,4,47,48</sup>.

**Structural connectivity analysis.** Structural connectivity was computed using eight regions as nodes and the fiber density between two regions as edges. The fiber density between two regions was defined as follows:  $fiber\ density(i, j) = \frac{2}{S_i + S_j} \sum_{f \in fibers} \frac{1}{l(f)} \times mean\ FA(i, j)$ , where  $f$  is the fiber connection between regions  $i$  and  $j$ ,  $l(f)$  is the length of the fiber connection,  $S_i$  is the surface area of region  $i$ , and  $mean\ FA(i, j)$  is the mean fractional anisotropy value along all fiber connections between regions  $i$  and  $j$ . In other words, fiber density was calculated as the connection efficiency between two regions, which was further defined as the number of fibers normalized by length and surface area<sup>49–53</sup>. The edge values were entered as elements of the matrix, referred to as the connectivity matrix. A simple network model that considered undirected and weighted edges was used to construct the brain network. We adopted DC, which is a simple and sensitive measure of structural connectivity, to measure the properties of the brain networks<sup>30</sup>. DC was computed as the sum of all edge weights connected in a given node<sup>54</sup>. DC represents the importance of a given node, where a high DC value signifies high information flow in the node. The microstructure of white matter is under strong genetic control, and thus, the white matter structural connectivity might be affected by genes<sup>55</sup>.

**Group-wise differences in neuroimaging.** Group-wise differences in neuroimaging were assessed with t-tests and Bonferroni correction between the HC and PD groups. For each group, the DC values of the eight nodes were stacked into three-dimensional matrices. Each group had a single three-dimensional matrix. Each element in the stacked connectivity matrix contained 40 observations. We performed t-tests to compare between the HC and PD groups to assess group-wise differences. Multiple comparison issues were corrected using the Bonferroni correction ( $p < 0.05$ , corrected)<sup>56,57</sup>. DC values of regions with significant differences were used as intermediate phenotypes for imaging genetics.

**Genetic quality control.** We quality controlled DNA genotyping data based on the ENIGMA protocol<sup>40</sup>. Quality control of genotype data was performed based on minor allele frequency, genotype missing rate, Hardy-Weinberg equilibrium, and genotyping rate. SNPs that did not satisfy the criteria were excluded, and subjects with a low genotyping rate were also excluded. In total, SNPs with a minor allele frequency <1%,

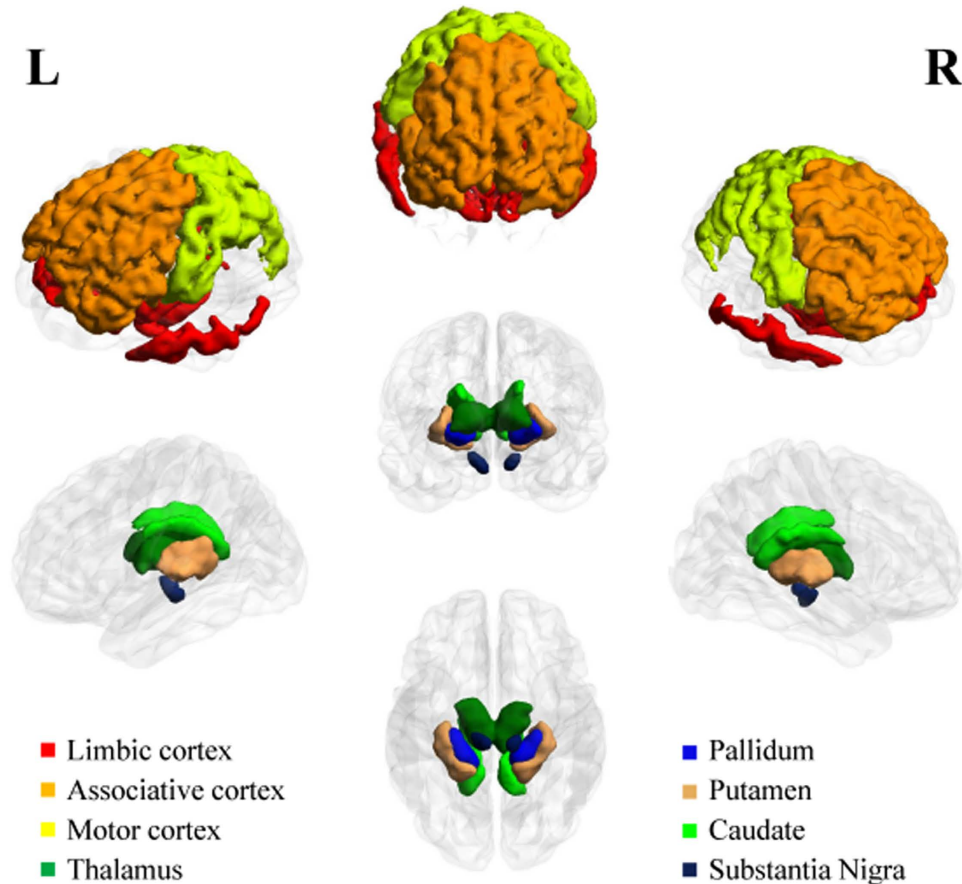


**Figure 2. Overview of neuroimaging and image genetics processing steps.**

Hardy-Weinberg equilibrium  $<10^{-6}$ , and a genotype missing rate  $>5\%$  were excluded from the study. Following quality control, the remaining SNPs were used for further analysis.

**Imaging genetics.** Imaging genetics analysis was conducted using intermediate phenotypes to identify genetic variants associated with PD. In the first analysis, neuroimaging features (i.e., DC values of regions with significant differences between HC and PD) were used as intermediate phenotypes. In the second analysis, the clinical score (i.e., MDS-UPDRS) was used as an intermediate phenotype. The genetic variance was described with an additive model<sup>58</sup>. Imaging genetics was performed using Plink software<sup>41</sup>. A linear regression analysis was performed at each SNP using the number of minor alleles as the independent variable and various types of intermediate phenotype information as the dependent variable. Age and sex were included as covariates in the linear model. Significant genetic variants were determined using a permutation test by performing random group assignments 10,000 times<sup>23,59</sup>. We focused on SNPs that were previously linked with PD.

**Construction of multiple linear regression models.** We constructed linear models using a stepwise approach to explain the clinical score (i.e., MDS-UPDRS) for all subjects (i.e., for both the HC and PD groups).



**Figure 3. ROI specifications.** Eight ROIs associated with PD in the CBGT circuit were defined based on MR images. The ROIs include the associative cortex, limbic cortex, sensorimotor cortex, caudate, putamen, pallidum, thalamus, and substantia nigra.

Multiple linear regression models were constructed using imaging features, genetic features, and two-way interactions between them. The multiple linear regression also considered family history information, MoCA, GDS, sex, age, and multi-center information as co-variants to control for confounding effects. Stepwise regression was applied to construct the regression models explaining the MDS-UPDRS scores<sup>60</sup>. An initial model with all possible independent variables along with interaction terms among them was constructed. Then, a candidate independent variable was removed if the removal of the variable led to a smaller p-value for the overall regression model. The process of variable removal was repeated until there was no improvement of the p-value of the regression model. For instance, the initial model considered DC values of four significant ROIs (from Table 1) and numerical SNP values of significant genetic variants (from Table 2) as independent variables. Then, each independent variable was tested for possible removal as described above. Subsequently, three linear regression models were constructed to explain the MDS-UPDRS score. The first model considered only neuroimaging features, the second model considered only genetic variants, and the third model considered both genetic variants and neuroimaging features:

$$Y = \alpha \times IMG + \beta \times COV + \gamma \times INTER + \varepsilon \quad (1)$$

$$Y = \alpha \times SNP + \beta \times COV + \gamma \times INTER + \varepsilon \quad (2)$$

$$Y = \alpha \times IMG + \beta \times SNP + \gamma \times COV + \delta \times INTER + \varepsilon, \quad (3)$$

where  $Y$  is the clinical score;  $IMG$  is the imaging features;  $SNP$  is the genetic features;  $COV$  is the co-variants;  $INTER$  are the two-way interactions of predictors;  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\delta$  are estimated coefficients; and  $\varepsilon$  is a residual error term. The quality of multiple linear regression was assessed with adj- $R^2$  values.

**Prediction and validation of the model.** To validate the models, a leave-one-out cross validation method was applied by assigning each case as the test set and the remaining cases as the training set. The process was repeated 80 times, each time assigning a different test set. The performance of prediction was assessed with Pearson's correlation coefficients between actual and predicted MDS-UPDRS scores. The RMSE was computed to quantify differences between the predicted and actual MDS-UPDRS.



## References

1. Stoessl, A. J. Neuroimaging in Parkinson's disease. *Neurotherapeutics* **8**, 72–81 (2011).
2. National Collaborating Centre for Chronic Conditions (Great Britain). Parkinson's disease: national clinical guideline for diagnosis and management in primary and secondary care. *Royal College of Physicians*, (2006).
3. Obeso, J. A. *et al.* Functional organization of the basal ganglia: therapeutic implications for Parkinson's disease. *Movement disorders: official journal of the Movement Disorder Society* **23** Suppl 3, S548–S559, doi: 10.1002/mds.22062 (2008).
4. Sharman, M. *et al.* Parkinson's disease patients show reduced cortical-subcortical sensorimotor connectivity. *Movement Disorders* **28**, 447–454 (2013).
5. Dauer, W. & Przedborski, S. Parkinson's disease: mechanisms and models. *Neuron* **39**, 889–909 (2003).
6. Rodriguez-Oroz, M. C. *et al.* Initial clinical manifestations of Parkinson's disease: features and pathophysiological mechanisms. *The Lancet Neurology* **8**, 1128–1139 (2009).
7. Ravina, B. *et al.* The role of radiotracer imaging in Parkinson disease. *Neurology* **64**, 208–215 (2005).
8. Wu, T. *et al.* Changes of functional connectivity of the motor network in the resting state in Parkinson's disease. *Neuroscience letters* **460**, 6–10 (2009).
9. Zhang, K. *et al.* Voxel-based analysis of diffusion tensor indices in the brain in patients with Parkinson's disease. *European journal of radiology* **77**, 269–273 (2011).
10. Zhang, Yu *et al.* Diffusion tensor imaging of the nigrostriatal fibers in Parkinson's disease. *Movement Disorders* **30**, 1229–1236 (2015).
11. Pavese, N. & Brooks, D. J. Imaging neurodegeneration in Parkinson's disease. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* **1792**, 722–729 (2009).
12. Le Bihan, D. *et al.* Diffusion tensor imaging: concepts and applications. *Journal of magnetic resonance imaging* **13**, 534–546 (2001).
13. Weingarten, C. P., Sundman, M. H., Hickey, P. & Chen, N.-k. Neuroimaging of Parkinson's disease: Expanding views. *Neuroscience & Biobehavioral Reviews* **59**, 16–52 (2015).
14. Khan, N. L. *et al.* Mutations in the gene LRRK2 encoding dardarin (PARK8) cause familial Parkinson's disease: clinical, pathological, olfactory and functional imaging and genetic data. *Brain* **128**, 2786–2796 (2005).
15. van Nuenen, B. F. *et al.* Mapping preclinical compensation in Parkinson's disease: an imaging genomics approach. *Movement Disorders* **24**, S703–S710 (2009).
16. Burton, P. R. *et al.* Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
17. Hirschhorn, J. N. & Daly, M. J. Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics* **6**, 95–108 (2005).
18. Klein, C. & Schlossmacher, M. G. The genetics of Parkinson disease: implications for neurological care. *Nature clinical practice Neurology* **2**, 136–146 (2006).
19. Potkin, S. G. *et al.* A genome-wide association study of schizophrenia using brain activation as a quantitative phenotype. *Schizophrenia bulletin* **35**, 96–108 (2009).
20. Thompson, P. M., Martin, N. G. & Wright, M. J. Imaging genomics. *Current opinion in neurology* **23**, 368 (2010).
21. Stein, J. L. *et al.* Voxelwise genome-wide association study (vGWAS). *neuroimage* **53**, 1160–1174 (2010).
22. Tost, H., Bilek, E. & Meyer-Lindenberg, A. Brain connectivity in psychiatric imaging genetics. *Neuroimage* **62**, 2250–2260 (2012).
23. Wang, Q. *et al.* Genome-wide association analysis with gray matter volume as a quantitative phenotype in first-episode treatment-naïve patients with schizophrenia. *PLoS one* **8**, e75083 (2013).
24. Marek, K. *et al.* The parkinson progression marker initiative (PPMI). *Progress in neurobiology* **95**, 629–635 (2011).
25. Disease, M. D. S. T. F. o. R. S. f. P. s. The Unified Parkinson's Disease Rating Scale (UPDRS): status and recommendations. *Movement disorders: official journal of the Movement Disorder Society* **18**, 738 (2003).
26. Kim, H. J. *et al.* Alterations of mean diffusivity in brain white matter and deep gray matter in Parkinson's disease. *Neuroscience letters* **550**, 64–68 (2013).
27. 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* **8**, e59717 (2013).
28. Van Dijk, K. R. *et al.* Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. *Journal of neurophysiology* **103**, 297–321 (2010).
29. Zuo, X.-N. *et al.* Network centrality in the human functional connectome. *Cerebral cortex* **22**, 1862–1875 (2012).
30. Rubinov, M. & Sporns, O. Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* **52**, 1059–1069 (2010).
31. Purves, D., Augustine, G. J., Fitzpatrick, D. *et al.* Neuroscience. 2nd edition. *Sunderland (MA): Sinauer Associates* (2001).
32. Ziegler, E. *et al.* Mapping track density changes in nigrostriatal and extranigral pathways in Parkinson's disease. *Neuroimage* **99**, 498–508 (2014).
33. Tolosa, E., Wenning, G. & Poewe, W. The diagnosis of Parkinson's disease. *The Lancet Neurology* **5**, 75–86 (2006).
34. Schlossmacher, M. G. *et al.* Parkin localizes to the Lewy bodies of Parkinson disease and dementia with Lewy bodies. *The American journal of pathology* **160**, 1655–1667 (2002).
35. Sriram, S. R. *et al.* Familial-associated mutations differentially disrupt the solubility, localization, binding and ubiquitination properties of parkin. *Human molecular genetics* **14**, 2571–2586 (2005).
36. Anders, S. *et al.* Compensatory premotor activity during affective face processing in subclinical carriers of a single mutant Parkin allele. *Brain* **135**, 1128–1140 (2012).
37. Nuytemans, K., Theuns, J., Cruts, M. & Van Broeckhoven, C. Genetic etiology of Parkinson disease associated with mutations in the SNCA, PARK2, PINK1, PARK7, and LRRK2 genes: a mutation update. *Human mutation* **31**, 763–780 (2010).
38. Park, H.-M. *et al.* The serine protease HtrA2/Omi cleaves Parkin and irreversibly inactivates its E3 ubiquitin ligase activity. *Biochemical and biophysical research communications* **387**, 537–542 (2009).
39. Cilenti, L. *et al.* Inactivation of Omi/HtrA2 protease leads to the deregulation of mitochondrial Mulan E3 ubiquitin ligase and increased mitophagy. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **1843**, 1295–1307 (2014).
40. ENIGMA2 Genetics Support Team. ENIGMA2 1KGP Cookbook (v3) [Online]. The Enhancing Neuroimaging Genetics through MetaAnalysis (ENIGMA) Consortium. [http://enigma.ini.usc.edu/wpcontent/uploads/2012/07/ENIGMA2\\_1KGP\\_cookbook\\_v3.pdf](http://enigma.ini.usc.edu/wpcontent/uploads/2012/07/ENIGMA2_1KGP_cookbook_v3.pdf) (2013).
41. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* **81**, 559–575 (2007).
42. Smith, S. M. *et al.* Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* **23** Suppl 1, S208–219, doi: 10.1016/j.neuroimage.2004.07.051 (2004).
43. Fischl, B. FreeSurfer. *Neuroimage* **62**, 774–781 (2012).
44. Wang, R. *et al.* Diffusion toolkit: a software package for diffusion imaging data processing and tractography. *Proc Intl Soc Mag Reson Med.* **5**, 3720 (2007).
45. Keuken, M. C. *et al.* Quantifying inter-individual anatomical variability in the subcortex using 7T structural MRI. *NeuroImage* **94**, 40–46 (2014).

46. Desikan, R. S. *et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968–980 (2006).
47. Parent, A. & Hazrati, L.-N. Functional anatomy of the basal ganglia. I. The cortico-basal ganglia-thalamo-cortical loop. *Brain Research Reviews* **20**, 91–127 (1995).
48. Braak, H. & Del Tredici, K. Cortico-basal ganglia-cortical circuitry in Parkinson's disease reconsidered. *Experimental neurology* **212**, 226–229 (2008).
49. Kim, M. & Park, H. Structural connectivity profile of scans without evidence of dopaminergic deficit (SWEDD) patients compared to normal controls and Parkinson's disease patients. *SpringerPlus* **5**, 1421 (2016).
50. Tuch, D. S., Wedeen, V. J., Dale, A. M., George, J. S. & Belliveau, J. W. Conductivity tensor mapping of the human brain using diffusion tensor MRI. *Proceedings of the National Academy of Sciences* **98**, 11697–11701 (2001).
51. Hagmann, P. *et al.* White matter maturation reshapes structural connectivity in the late developing human brain. *Proceedings of the National Academy of Sciences* **107**, 19067–19072 (2010).
52. Fischi-Gómez, E. *et al.* Structural brain connectivity in school-age preterm infants provides evidence for impaired networks relevant for higher order cognitive skills and social cognition. *Cerebral cortex* **25**, 2793–2805 (2015).
53. Feldman, H. M., Yeatman, J. D., Lee, E. S., Barde, L. H. & Gaman-Bean, S. Diffusion tensor imaging: a review for pediatric researchers and clinicians. *Journal of developmental and behavioral pediatrics: JDBP* **31**, 346 (2010).
54. Schwarz, A. J. & McGonigle, J. Negative edges and soft thresholding in complex network analysis of resting state functional connectivity data. *Neuroimage* **55**, 1132–1146 (2011).
55. Thompson, P. M., Ge, T., Glahn, D. C., Jahanshad, N. & Nichols, T. E. Genetics of the connectome. *Neuroimage* **80**, 475–488 (2013).
56. Nichols, T. E. & Holmes, A. P. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human brain mapping* **15**, 1–25 (2002).
57. Bullmore, E. T. *et al.* Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. *Medical Imaging, IEEE Transactions on* **18**, 32–42 (1999).
58. Brown, G. G. *et al.* Voxel-based morphometry of patients with schizophrenia or bipolar I disorder: a matched control study. *Psychiatry Research: Neuroimaging* **194**, 149–156 (2011).
59. Smith, S. M. *et al.* Functional connectomics from resting-state fMRI. *Trends in cognitive sciences* **17**, 666–682 (2013).
60. Draper, N. R. & Smith, H. *Applied regression analysis.* (John Wiley & Sons, 2014).

## Acknowledgements

This study was supported by the Institute for Basic Science (grant number IBS-R015-D1). This work was also supported by National Research Foundation of Korea (grant number NRF-2016R1A2B4008545). Imaging data were obtained from the PPMI. PPMI is a public-private partnership funded by The Michael J. Fox Foundation for Parkinson's Research, Abbott, Avid Pharmaceuticals, Biogen Idec, Bristol-Myers Squibb, Covance, Elan, GE Healthcare, Genentech, GlaxoSmithKline, Lilly, Merck, Meso Scale Discovery, Pfizer, Roche, and UCB ([www.ppmi-info.org/fundingpartners](http://www.ppmi-info.org/fundingpartners)).

## Author Contributions

H.P. and M.K. conceptualized the design of the study. M.K., J.K., and S.L. performed the experiments and analyses. H.P. oversaw the entire study. H.P. and M.K. co-wrote the draft. All authors reviewed and provided feedback for the draft.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing Interests:** The authors declare no competing financial interests.

**How to cite this article:** Kim, M. *et al.* Imaging genetics approach to Parkinson's disease and its correlation with clinical score. *Sci. Rep.* **7**, 46700; doi: 10.1038/srep46700 (2017).

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017