Learning Phenotypic Associations for Parkinson's Disease with Longitudinal Clinical Records

Weishen Pan, PhD[1], Chang Su, PhD[1], Jacqueline R. M. A. Maasch, MSE[2], Kun Chen, PhD[3], Claire Henchcliffe, MD, DPhil[4], Fei Wang, PhD[1,*]

[1]Department of Population Health Sciences, Weill Cornell Medicine, Cornell University, New York, NY, USA; [2]Department of Computer Science, Cornell Tech, Cornell University, New York, NY, USA; [3]Department of Statistics, University of Connecticut, Storrs, CT, USA; [4]Department of Neurology, University of California, Irvine, Irvine, CA, USA

Abstract

Parkinson's disease (PD) is associated with multiple clinical motor and non-motor manifestations. Understanding of PD etiologies has been informed by a growing number of genetic mutations and various fluid-based and brain imaging biomarkers. However, the mechanisms underlying its varied phenotypic features remain elusive. The present work introduces a data-driven approach for generating phenotypic association graphs for PD cohorts. Data collected by the Parkinson's Progression Markers Initiative (PPMI), the Parkinson's Disease Biomarkers Program (PDBP), and the Fox Investigation for New Discovery of Biomarkers (BioFIND) were analyzed by this approach to identify heterogeneous and longitudinal phenotypic associations that may provide insight into the pathology of this complex disease. Findings based on the phenotypic association graphs could improve understanding of longitudinal PD pathologies and how these relate to patient symptomology.

Introduction

Multiple identifiable and quantifiable phenotypic features are associated with Parkinson's disease (PD), including motor manifestations (such as bradykinesia, muscle rigidity, tremor, and postural instability), non-motor manifestations (such as depression, cognitive decline, fatigue, and dysautonomia), and biomarkers. How these features are linked remains to be precisely understood, and heterogeneity among individual patients diagnosed with PD leads to further complexity. However, these features are related to the spatially, temporally, and molecularly complex pathologies of PD, and may therefore provide insight into underlying networks that are affected as PD progresses. There is therefore an urgent need to identify relationships among these heterogeneous phenotypic features to further uncover the underlying mechanisms of PD. Most existing cohort studies focus on a single specific feature and its relationships to other factors relying on statistical testing, univariate regression, or multivariate regression[1-4]. However, these approaches require hypotheses of the independent and dependent features to be tested, and hence may not be appropriate for detecting the complex correlations among the heterogeneous features of PD (Figure 1). Moreover, traditional univariate/multivariate regression methods require assumptions about the functional form of the relationships between variables (e.g., linearity, polynomiality).

To address these limitations, we present a data-driven approach based on the PC (Peter-Clark) algorithm[5]. This algorithm takes all phenotypic features as inputs and produces an undirected graph, i.e., the phenotypic association graph, revealing the subtle relationships among the features (Figure 1). For cross-cohort validation, the phenotypic association graphs within three PD cohorts were computed. Moreover, by investigating the phenotypic association graphs generated at different stages of PD, we analyzed how feature relationships evolve as PD progresses. Finally, the obtained phenotypic association graphs are used to discuss the complex pathologies of PD. The supplemental materials are available at https://github.com/weishenpan15/pd-association-graph.

Methods

Study population. Publicly available data from two cohorts comprising individuals with PD and healthy control (HC) subjects were obtained: the Parkinson's Progression Markers Initiative (PPMI) and the Parkinson's Disease Biomarkers Program (PDBP). The Fox Investigation for New Discovery of Biomarkers (BioFIND) provides data for one additional cohort comprised of individuals with PD and no controls[6-8].

The PPMI study is a prospective longitudinal study of de novo PD patients who were untreated with medications at baseline and were enrolled at 33 sites internationally[6]. The institutional review board of the University of Rochester

*Corresponding author. Email: few2001@med.cornell.edu

(NY, USA) and each PPMI participating site approved the PPMI study protocol. Data downloaded from https://www.ppmi-info.org/data on July 01, 2020 under PPMI Data Use Agreement, were used for this analysis. At that time, enrollment was complete. 424 PD patients with available baseline data were included for analysis. Baseline and follow-up assessments included complete neuropsychological test data at 1-, 2-, 3-, 4-, and 5-year follow-up.

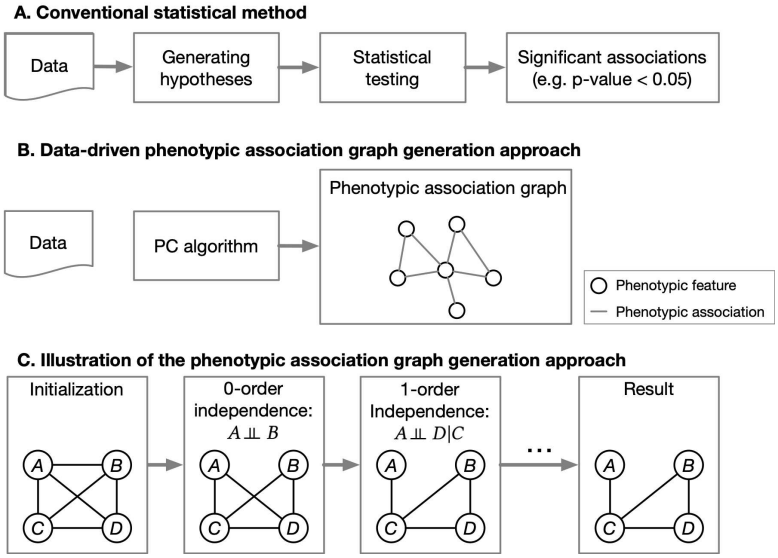

Figure 1: An illustration of conventional phenotypic association identification methods (A), the proposed method for data-driven phenotypic association graph generation (B) and the illustration of the method (C).

The PDBP study is a longitudinal study of PD patients with different severities at baseline who were enrolled at 11 US sites[7]. The institutional review board of each PDBP participating site approved the study protocol for that site. Data of PDBP were downloaded via the Accelerating Medicines Partnership Parkinson's disease (AMP-PD) platform (http://amp-pd.org) on Feb 01, 2020 under AMP-PD Data Use Agreement. Enrollment was complete at that time. A total of 836 PD patients with available baseline data were included for analysis. Patients' data at baseline and 1- and 2-year follow-up were used for analysis.

The BioFIND study is a cross-sectional cohort of participants enrolled at 8 sites in the United States[8]. The average duration of PD in this cohort is 8.34 years, which is much longer than the PD duration of subjects in the PPMI, and enrolled patients were receiving symptomatic treatment. The institutional review board of BioFIND approved the study protocol. Data downloaded from https://biofind.loni.usc.edu on July 01, 2020 under BioFIND Data Use Agreement, were used for this analysis. A total of 126 moderate-advanced PD participants were included for analysis.

Table 1. Summary of benchmark data sets.

	Dataset and sample selected	**Features**
Setting1	PPMI, PD at baseline before symptomatic treatment, 1–5 years follow-up	Features that are shared in all periods of PPMI
Setting 2	BioFIND, PD	Features that are shared among BioFIND, PDBP, and PPMI longitudinally
Setting 3	PDBP, PD at baseline, 1–2 years follow-up	Features that are shared among BioFIND, PDBP, and PPMI longitudinally
Setting 4	PPMI, PD at baseline before symptomatic treatment	Features included at baseline of PPMI

Phenotypic features. A wide range of phenotypic features were included for analysis, including motor and non-motor manifestations based upon validated rating scales, CSF biomarkers, and MRI neuroimaging data. In addition, other

features including demographics, genetic risk score, and medications were included. The details of all features used for analysis are listed in Tables S1 and S2 of the supplementary material, in which we have matched features collected among PPMI, PDBP, and BioFIND when possible. Specifically, we used sub-scores of the Movement Disorder Society Unified Parkinson Disease Rating Scale (MDS-UPDRS) Part I, extracting responses to individual questions regarding fatigue, hallucinations, and apathy[9]. We also analyzed composite scores for tremor and postural instability and gait disturbance (PIGD) as motor features.

Phenotypic association graph identification. The PC algorithm was used to generate the association graphs of selected PD features[5]. This algorithm starts from a fully connected graph, of which the nodes are all features studied and the undirected edges among them are candidate associations. The algorithm then determines which edges should be deleted by conditional independence testing. When the algorithm converges, we obtain the phenotypic association graph of the clinical features. An example with 4 variables is illustrated in Figure 1(C). A full description of the algorithm is shown in the supplementary materials.

To compare the associations among features in different cohorts (PPMI, PDBP, and BioFIND) and PD durations (within the PPMI cohort), we ran the PC algorithm upon different combinations of datasets and features following the settings in Table 1. We first ran the PC algorithm on data from PPMI subjects at different time points to identify dynamic relationships among features (Setting 1). For cross-cohort comparison, we ran the algorithm on PDBP and BioFIND data with features shared in all three cohorts (Settings 2 and 3). Features included in each cohort are listed in Supplementary Tables S1 and S2. The cerebrospinal fluid (CSF) biomarkers (Aβ1-42, total tau (t-tau), phosphorylated tau (p-tau), α-synuclein) and brain imaging data (magnetic resonance imaging (MRI), single photon emission computed tomography using 131I-ioflupane (DaTscan])) are available in both PPMI-baseline. We ran the algorithm with Setting 4 in Table 1. For MRI features, we used surface area, cortical volume, and white matter volume from 34 brain regions defined by the Desikan-Killiany atlas, which is listed in Table S3 in supplementary materials.

Feature grouping based on graph structure. When generating phenotypic association graphs with longitudinal data from the PPMI study, feature-level analyses are too fine-grained to overcome noise. For example, two features may be linked directly at baseline and 2 years after baseline, while the edge may disappear at 1-year follow-up. This is due to the limited number of subjects in the studied cohorts. To address this issue, phenotypic features were partitioned into different groups according to graph structure. This partitioning process facilitates the analysis of group-level relationships, which are more consistent than feature-level associations. Specifically, we first constructed a weighted summarization graph in which nodes are all clinical features (excluding demographics, medicine, biomarkers, and brain imaging features) and in which the weight of each connection is the frequency of occurrence of the connection over all periods. The Louvain community detection algorithm[10], which aims at clustering nodes into densely intra-connected yet sparsely inter-connected groups, was then utilized to group the features on the summarization graph (described in more detail in the appendix). Some groups obtained by the community detection algorithm were further separated into sub-groups consisting of clinical features from finer domains.

Results

Phenotypic association graphs from PPMI. The phenotypic association graphs from PPMI were generated based on Setting 1 in Table 1. Figure 2 illustrates the generated graph at baseline. Nine feature groups were identified, including Motor, Mood, Cognitive-Hopkin's Verbal Learning Test (HVLT), Cognitive-Montreal Cognitive Assessment (MoCA), Cognitive-Executive, Cognitive-Visuospatial, Autonomic Dysfunction & Sleep Disorder (ADSD), Hallucinations, and Impulse Control Disorder (ICD) (Table 2). We observed that clinical features in similar medical domains are grouped together. For example, MDS-T (tremor score), MDS-P (postural instability and gait disturbance score), and Hoehn and Yahr (H&Y) stage were grouped under Motor, while Geriatric Depression Scale (GDS), State-Trait Anxiety Inventory (STAI), MDS-apathy, and MDS-fatigue were grouped under Mood. Fine-grained (i.e., feature-level) graphs at 1- to 5-year follow-up of PPMI are shown in Supplemental Figures S1-S5. If two features are independent of each other conditioned on other features, they will not be directly connected. In other words, an edge linking two features indicates that they are always dependent conditioned on any set of other features.

By considering each feature group as a super-node, we generated the group-level association graphs of different periods in PPMI. As shown in Figure 3, connections between different groups are relatively sparse at baseline while more connections between groups emerge as PD progresses. To interpret these findings, we categorized the edges between features into two groups: (i) the stable edges occurring throughout the different time periods examined, and (ii) the longitudinal edges that change as the disease progresses.

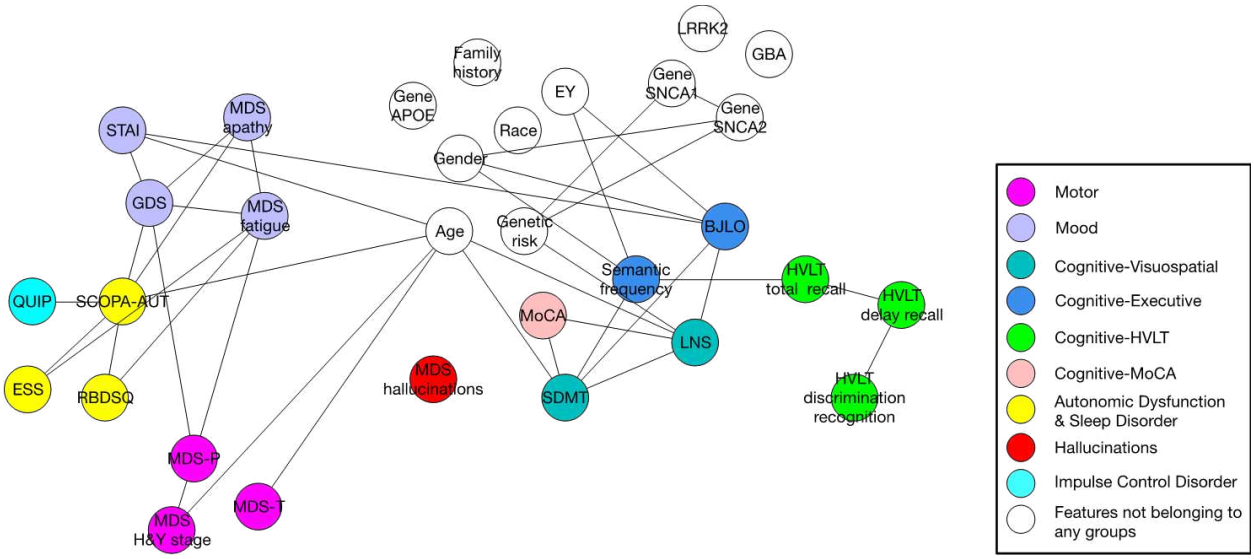

Figure 2. The feature-level phenotypic association graph for the PPMI cohort at baseline. The selected samples and features follow Setting 1 in Table 1. Each node represents a feature, while node color represents the group to which a specific feature belongs. Nodes in white are those features that do not participate in any phenotypic grouping. Abbreviation: MDS=Movement Disorder Society. MDS-T=MDS Tremor score. MDS-P=MDS Postural Instability and Gait Difficulty score. H&Y=Hoehn and Yahr. GDS=Geriatric Depression Scale. STAI=State Trait Anxiety Inventory. HVLT=Hopkin's Verbal Learning Test. MoCA=Montreal Cognitive Assessment. SDMT=Symbol Digit Modalities Test. LNS=Letter Number Sequencing. BJLO=Benton Judgment of Line Orientation. QUIP=Questionnaire for Impulsive-Compulsive Disorders in Parkinson's Disease. SCOPA-AUT=SCales for Outcomes in PArkinson's disease-AUTomotic symptoms. ESS=Epworth Sleepiness Scale. RBDSQ=REM Sleep Behavior Disorder Screening Questionnaire.

Table 2. List of the feature groups obtained from the phenotypic association graph of PPMI cohort.

Group	**Features**
Motor	MDS-T, MDS-P, H & Y stage
Mood	GDS, STAI, MDS-apathy, MDS-fatigue
Cognitive-HVLT	HVLT total recall, HVLT delayed recall, HVLT discrimination, recognition
Cognitive-MoCA	MoCA
Cognitive-Executive	Semantic Frequency, BJLO
Cognitive-Visuospatial	SDMT, LNS
Autonomic Dysfunction & Sleep Disorder (ADSD)	SCOPA-AUT, ESS, RBDSQ
Impulse Control Disorder	QUIP
Hallucinations	MDS-hallucinations

Phenotypic association graph from PDBP and BioFIND. Phenotypic association graphs from the PDBP and BioFIND cohort data were generated based on Settings 2 and 3 in Table 1. Five shared feature groups were included, including Motor, Mood, Cognitive-MoCA, ADSD, and Hallucinations. The group-level association graphs are shown in Figure 4. The feature-level association graphs at baseline of PDBP are shown in Figure 5, while those at 1- and 2-year follow-up are shown in Supplemental Figures S6 and S7. The feature-level association graph from BioFIND is shown in Figure 6.

Except for feature groups of PPMI that are not available in the PDBP and BioFIND cohorts, the graphs derived from PDBP and BioFIND (Figure 4) are similar to those generated from the later-period data of PPMI (Figure 3). We also observed similar results on feature-level graphs (Figures 5, 6, and Supplementary Figures S6, S7), as PDBP and BioFIND subjects have a longer duration of PD at enrollment compared to PPMI.

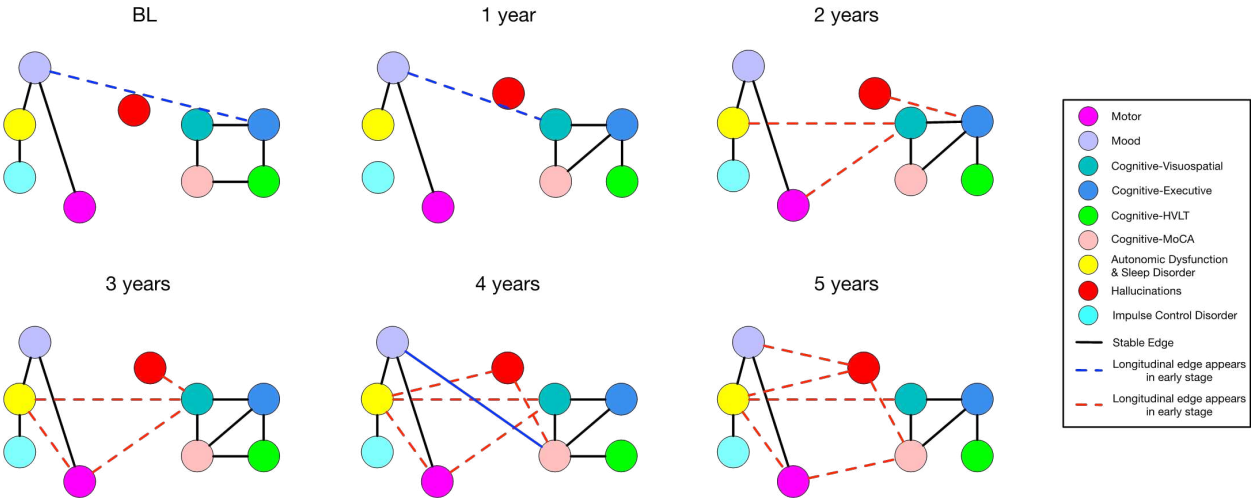

Figure 3. Group-level longitudinal phenotypic association graphs for the PPMI cohort. The selected samples and features follow setting 1 in Table 1. Each node represents an extracted feature group. Stable edges (i.e., those that appear in all time points) are marked as solid black lines. Longitudinal edges are marked in dashed blue (i.e., those that appear in the early stage) or in dashed red (i.e., those that appear in the late stage). HVLT=Hopkin's Verbal Learning Test. MoCA=Montreal Cognitive Assessment.

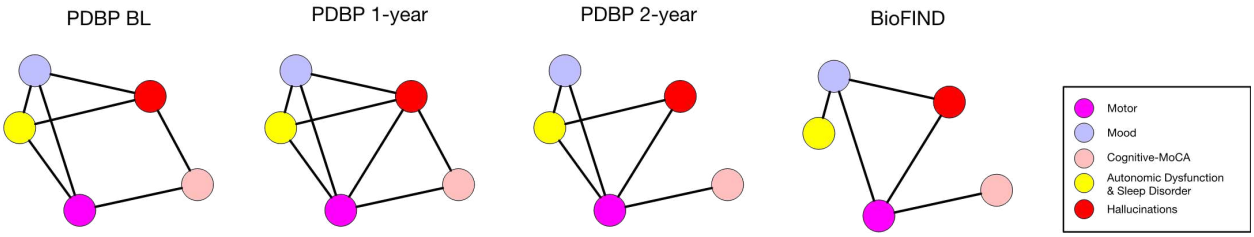

Figure 4. Group-level phenotypic association graphs for the PDBP and BioFIND cohorts. The selected samples and features follow Setting 3 in Table 1. Each node represents an extracted feature group. MoCA=Montreal Cognitive Assessment.

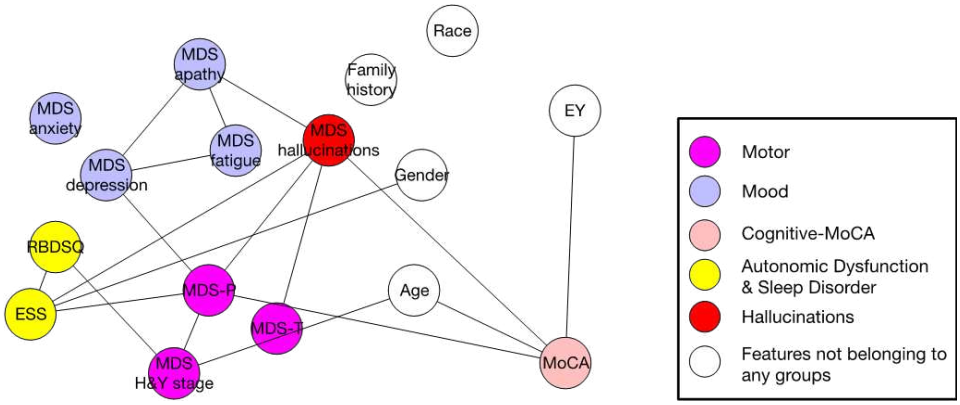

Figure 5. The feature-level phenotypic association graph for the PDBP cohort at baseline. MoCA=Montreal Cognitive Assessment.

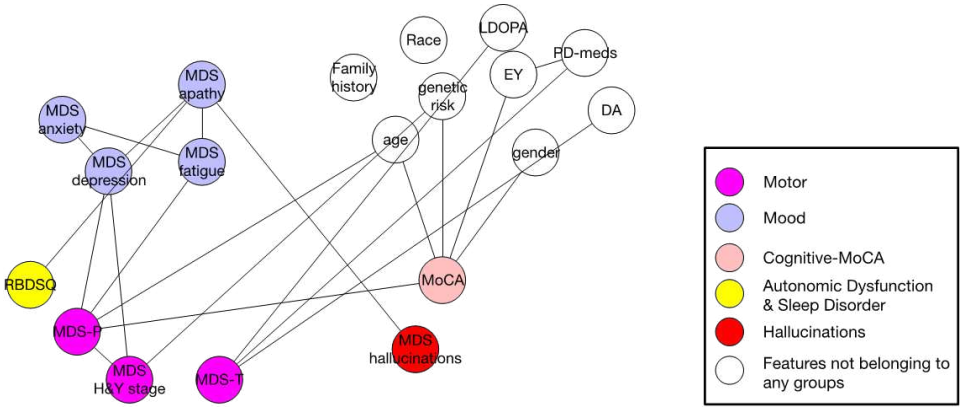

Figure 6. The feature-level phenotypic association graph for the BioFIND cohort. MoCA=Montreal Cognitive Assessment.

Phenotypic association graphs with CSF biomarkers and MRI features. To further investigate the correlations between CSF biomarkers and clinical features, we incorporated CSF biomarkers into our approach and generated the phenotypic association graphs of PPMI baseline data according to Setting 4 in Table 1. The resulting association graphs are shown in Figure 7, including only the MRI features which are directly connected with clinical characteristics.

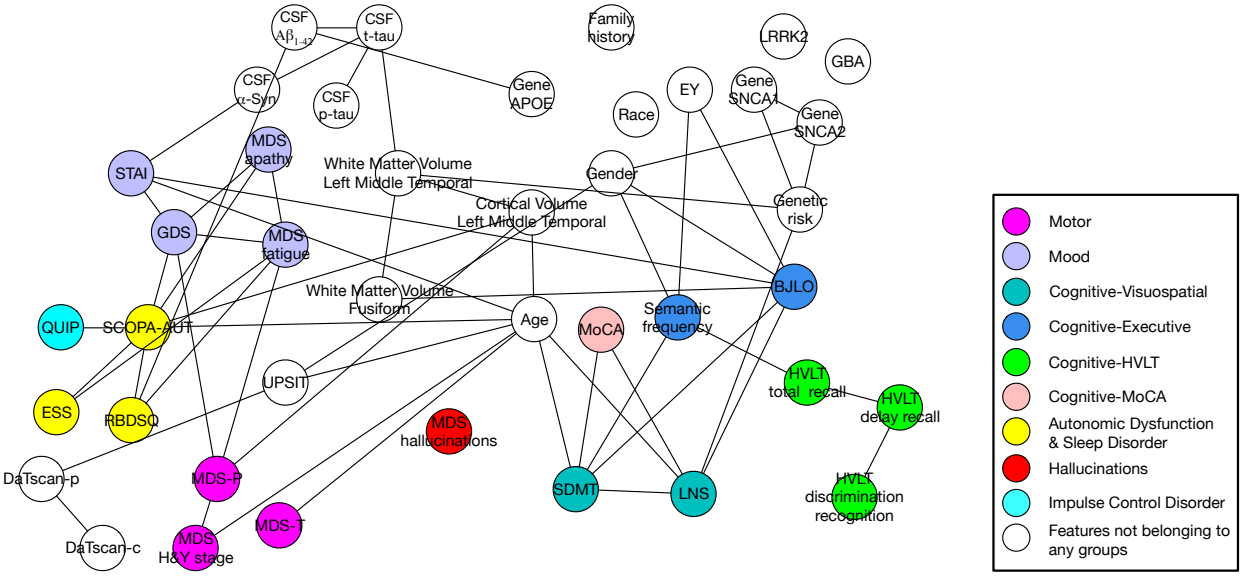

Figure 7. The feature-level phenotypic association graph at baseline of PPMI cohort, after incorporating CSF biomarkers and brain imaging.

Discussion

Stable phenotypic associations in PD. From the graphs shown in Figures 2-6 and Supplementary Figures S1-S7, we identified the following stable phenotypic associations.

(1) Stable associations were observed among Cognitive-Executive, Cognitive-Visuospatial, Cognitive-Hopkin's Verbal Learning Test (HVLT), and Cognitive-Montreal Cognitive Assessment (MoCA) groups. This is consistent with the fact that features in the four groups all serve as measures of cognitive manifestations.

(2) Stable associations were observed among features in the Mood group (including depression, anxiety, apathy, and fatigue), and feature-level intra-group connections such as the associations between Geriatric Depression Scale (GDS) and State-Trait Anxiety Inventory (STAI), between GDS and MDS-apathy, and between MDS-apathy and MDS-fatigue. The potential reasons could be two-fold: on one hand, the measures of depression, anxiety, apathy, and

fatigue have overlapping content in their respective scale questionnaires; on the other hand, these PD symptoms likely overlap in pathology through association with the production of dopamine, serotonin, and other neurochemicals[11-13].

(3) Stable associations were identified between Mood and Autonomic Dysfunction & Sleep Disorder (ADSD) features. These two groups contain common non-motor neurodegenerative features of PD: depression (GDS), anxiety (STAI), REM sleep behavior disorder (REM Sleep Behavior Disorder Screening Questionnaire, RBDSQ), excessive daytime sleepiness (Epworth Sleepiness Scale, ESS), and autonomic dysfunction (SCales for Outcomes in PArkinson's disease-AUTomotic symptoms, SCOPA-AUT). These features have a critical impact on the patient's health‐related quality of life. Besides the group-level connections, there exist consistent feature-level connections, such as associations between GDS and SCOPA-AUT, ESS and SCOPA-AUT, as well as RBDSQ and SCOPA-AUT. Previous studies have reported that these features are correlated with the severity of PD and are mostly correlated with each other[1-4]. Based on the PPMI data, any two of these features are significantly associated with each other (p-value < 0.01). These findings could potentially be explained by the spread of neurodegeneration within the brainstem. Braak et al reported that Lewy pathology and neuronal loss have been identified in the locus coeruleus, raphe nuclei, dorsal motor nucleus of the vagus (DMV), and pedunculopontine nuclei (PPN) as PD processes[14]. These regions are implicated in the control of sleep, mood, and autonomic function[3,4,15]. Previous pathological studies have also supported such associations[16,17]. Our results add strength to the evidence that lesions in these regions can cause sleep, mood, and autonomic dysfunctions. In addition, we also detected stable associations between ESS and RBDSQ within the ADSD group, which could be explained by similar underlying pathologies[18].

(4) Stable associations between mood and motor features were also observed. At the feature level, there are connections between MDS-P (MDS Postural Instability and Gait Difficulty score) and GDS. PD patients of the postural instability and gait difficulty (PIGD) subtype are more likely to have depression than those of the tremor-dominant subtype[19,20]. In addition, the relationship between GDS and MDS-P could be explained by a common pathology affecting the locus coeruleus, since this structure is implicated in the control of mood and in features of the PIGD subtype (e.g., postural stability)[3,21].

(5) The use of anti-PD medications is linked to features in the ADSD group. Though feature-level connections may not be very consistent, results suggest that ESS is connected to PD medication. This could potentially be explained by the side effect profile of PD medications, as they may lead to orthostatic hypotension, excessive daytime drowsiness, and sleep disruption[22].

Longitudinal phenotypic associations in PD. Based on the phenotypic association graphs generated at different time periods within PPMI, we identified associations that vary over time.

(1) An association between the features of the Motor, Cognitive-Visuospatial, and Cognitive-MoCA groups begins to occur at 2–5 years after baseline (Figure 3). At the feature level, there is an edge connecting MDS-P and Symbol Digit Modalities Test (SDMT)/MoCA at 2–5 years after baseline (Supplementary Figures S2-S5). This suggests that a common neurodegenerative process may cause cognitive dysfunction (especially processing speed and attentional dysfunction measured by SDMT) and the PIGD phenotype. Such neurodegenerative processes may include degeneration of dopaminergic systems affecting parallel basal ganglia-thalamocortical pathways and non-dopaminergic degeneration within cholinergic systems[23,24].

(2) Associations connecting motor symptoms to ADSD symptoms emerge at 3 years after baseline (Figure 3). At the feature level, there are connections between SCOPA-AUT and MDS-P (Supplementary Figures S3-S5). According to previous clinical studies, non-motor features comprising sleep and autonomic dysfunction are particularly related to the PIGD phenotype. The edge between SCOPA-AUT and MDS-P is likely due to degeneration within the noradrenergic and cholinergic systems. As noted above, the PIGD phenotype is associated with cholinergic degeneration and autonomic dysfunction in PD is also linked to cholinergic pathways that may include structures such as the DMV and PPN[25]. Connections between SCOPA-AUT and MDS-P suggest the existence of multiple non-dopaminergic pathologies that are related to the noradrenergic and cholinergic systems. Although sleep problems (e.g., REM sleep behavior disorder and excessive daytime sleepiness) are also found to be correlated with the PIGD phenotype, RBDSQ, and ESS are individually independent of MDS-P and conditioned on other variables such as SCOPA-AUT. This suggests that there is likely no shared pathway between sleep problems and the PIGD phenotype. The correlation between them arises from parallel pathways, some of which affect autonomic functions and sleep and some of which affect autonomic functions and PIGD.

(3) Connections between the ADSD and Cognitive-Visuospatial groups were observed in the later periods analyzed (2–5 years after baseline), especially between SCOPA-AUT and SDMT. This finding is consistent with existing

clinical studies that reported correlations between autonomic dysfunction and cognitive impairment in advanced PD but not in early-stage PD patients[26,27].

(4) There are relationships between hallucinations and ADSD, and Cognitive-relative groups in relatively late-stage patients. These associations include feature-level relations between hallucinations and SCOPA-AUT and between hallucinations and MoCA/SDMT/Benton Judgment of Line Orientation (BJLO). As for the connection between hallucinations and cognitive impairment, multiple structural and functional neuroimaging studies have reported overlap in brain region abnormalities in PD patients with hallucinations and those with cognitive impairment[28]. The correlation between autonomic dysfunction and hallucinations was also reported in earlier studies[29,30]. Although some previous studies explain severe autonomic dysfunction as reflecting more advanced PD[29], our results suggest that there may exist a more direct pathology affecting both autonomic dysfunction and hallucination. This may include pathology of the DMV and other brainstem nuclei, in addition to dysfunction of peripheral autonomic nerves[30]. Moreover, the connection between hallucinations and autonomic dysfunction is likely to be at least a partial consequence of cholinergic degeneration.

(5) A connection between the hallucinations and apathy groups is observed at the last time point in this study (5 years after baseline). Though this correlation was reported in previous studies[31], a lack of clear explanatory factors warrants further investigation.

(6) Although genetic risk score is a good predictor of PD, there is no significant correlation between genetic risk score and motor or non-motor features. Though genetic risk score is linked to cognitive features in the early stage of PD, this correlation disappears in more advanced periods (3–5 years after baseline). This might suggest that genetic risk score is associated with cognitive function[32]. However, other non-genetic factors affecting cognition arise as the disease progresses, obscuring the genetic effect such that it is not significant enough to be detected.

(7) An inconsistent association exists between the Mood group at baseline and the Cognitive-relative group at 1 and 4 years after baseline. In the early stage, there exist feature-level connections between STAI and BJLO/Letter Number Sequencing (LNS), while in the later stage, the connection between GDS and MoCA emerges. The correlation between depression/anxiety and cognitive impairment has been frequently reported[33]. One possible mechanism underlying this relationship is striatal dopamine degeneration[34]. Another study reported an association between cognitive impairment and both depression and cholinergic deficits, suggesting that cholinergic deficits could underlie depression. Consistent with our findings are reported associations between cortical cholinergic denervation and depression in PD patients, independent of their cognitive functioning[35]. In particular, the connection between GDS and MoCA appears 4 years after baseline, coincident in time with the triangle connections among SCOPA-AUT, MDS-P, and MoCA/SDMT. Due to its early-stage emergence, the relation between anxiety and cognitive impairment suggests that these symptoms may arise from different pathways. However, this hypothesis merits further investigation.

Cross-cohort comparison between PPMI, PDBP, and BioFIND cohorts. As shown in Figure 4 and Supplementary Figures S4-S5, the phenotypic association graph in BioFIND is similar to those graphs of PPMI at the later time points. For PDBP and BioFIND, there are edges between age/gender/years of education and cognitive function (MoCA), between MDS-depression and MDS-anxiety, between MDS-depression and MDS-apathy, between MDS- depression and MDS-fatigue, between MDS-depression and MDS-P, and between MDS-apathy and MDS-hallucinations. All of these edges are also presented as stable edges in PPMI. One inconsistent finding is that MDS-depression and RBDSQ are independent conditioned on MDS-apathy in BioFIND, while GDS and RBDSQ should be dependent conditioned on MDS-apathy with SCOPA-AUT excluded. If GDS is replaced with MDS-depression in PPMI, MDS-depression and RBDSQ are also found to be independent conditioned on MDS-apathy. This indicates that MDS-depression as measured with a single question may not be reliable relative to GDS. In addition, there are connections between MDS-P and MoCA in PDBP, consistent with our findings within PPMI. Our results are therefore consistent, as BioFIND includes PD patients with more advanced disease.

Phenotypic associations between CSF biomarkers, MRI features, and clinical features. As shown in Figure 7, we identified new associations after incorporating CSF-biomarker and brain imaging features for analysis in the PPMI dataset. Besides strong connections among different CSF biomarkers, there are relations connecting CSF biomarkers with RBDSQ and STAI. The relation between anxiety and CSF biomarkers has also been reported in previous work on PPMI and an Alzheimer's disease cohort[36,37]. This suggests that there is potentially a direct effect of α-synuclein pathology on particular areas of the brain or that those neurochemical deficits (such as neurotransmitter deficits) are involved in anxiety[36]. The connection between CSF-Aβ1-42 and RBDSQ may indicate a potential β-Amyloid pathway causing RBD, which has not been well-studied and merits discussion. Moreover, the path connecting the APOE gene, CSF-Aβ1-42, and RBDSQ suggests that APOE can be the genetic factor causing the RBD problem.

We also observed a connection between DaTscan-p (DaTscan Putamen) and H&Y stage. This is consistent with a role for DaTscan not only in diagnosis but also as a marker of severity of PD. DaTscan and olfaction are both predictors of PD, and the connection in our study between UPSIT and DaTscan-p suggests that dopamine transporter deficit and olfactory dysfunction are not independent factors of PD and they may be related by a complex pathway, or they may both act as surrogates for pathophysiological effects on dopaminergic cells[38].

From Figure 7, we also observed MRI features from brain regions of the left middle temporal and fusiform as important mediators between age/genetic risk scores and clinical features. These regions have been found to be associated with PD or PD-related clinical features[39,40]. Our findings show that they directly affect SCOPA-AUT, MDS-P, and BJLO. Considering the connections between these features when we analyze the stable/longitudinal edges, the regions of the left middle temporal and fusiform region could be involved in the pathologies of these features.

Conclusion

This paper describes PD phenotypic associations using data from the PPMI, PDBP, and BioFIND cohorts. The proposed methodology presents an algorithm for data-driven feature association graph construction. Applying our approach to patient data at baseline and 1-, 2-, 3-, 4-, and 5-year follow-up of the PPMI cohort, we identify patterns of phenotypic associations indicating the clinical progression of PD. Through cross-cohort comparison, similar graph structures were observed for PDBP, BioFIND, and late-period (4- and 5-year follow-up) data from PPMI. We suggest that the associations detected provide insight into the spread of neuropathology and networks of affected structures, and that the presented methodology can support hypothesis generation for clinical data mining.

Acknowledgment

This work was supported by MJFF 14858.01 and NSF 1750326.

References

1. Neikrug AB, Maglione JE, Liu L, Natarajan L, Avanzino JA, Corey-Bloom J, et al. Effects of Sleep Disorders on the Non-Motor Symptoms of Parkinson Disease. J Clin Sleep Med. 2013;09(11):1119-29.
2. Arnao V, Cinturino A, Valentino F, Perini V, Mastrilli S, Bellavia G, et al. In patient's with Parkinson disease, autonomic symptoms are frequent and associated with other non-motor symptoms. Clin Auton Res. 2015;25(5):301-7.
3. Simuni T, Caspell-Garcia C, Coffey C, Chahine LM, Lasch S, Oertel WH, et al. Correlates of excessive daytime sleepiness in de novo Parkinson's disease: A case control study. Mov Disord. 2015;30(10):1371-81.
4. Cui SS, Du JJ, Fu R, Lin YQ, Huang P, He YC, et al. Prevalence and risk factors for depression and anxiety in Chinese patients with Parkinson disease. BMC Geriatr. 2017;17(1):1-10.
5. Spirtes P, Glymour CN, Scheines R. Causation, prediction, and search. MIT press; 2000.
6. Marek K, Jennings D, Lasch S, Siderowf A, Tanner C, Simuni T, et al. The Parkinson Progression Marker Initiative (PPMI). Prog Neurobiol. 2011;95(4):629-35.
7. Rosenthal LS, Drake D, Alcalay RN, Babcock D, Bowman FD, Chen-Plotkin A, et al. The NINDS Parkinson's disease biomarkers program. Mov Disord. 2015;31(6):915-23.
8. Kang UJ, Goldman JG, Alcalay RN, Xie T, Tuite P, Henchcliffe C, et al. The BioFIND study: Characteristics of a clinically typical Parkinson's disease biomarker cohort. Mov Disord. 2016;31(6):924-32.
9. Goetz CG, Tilley BC, Shaftman SR, Stebbins GT, Fahn S, Martinez-Martin P, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results. Mov Disord. 2008;23(15):2129-70.
10. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech. 2008;2008(10):P10008.
11. Pavese N, Metta V, Bose SK, Chaudhuri KR, Brooks DJ. Fatigue in Parkinson's disease is linked to striatal and limbic serotonergic dysfunction. Brain. 2010;133(11):3434-43.
12. Yamanishi T, Tachibana H, Oguru M, Matsui K, Toda K, Okuda B, et al. Anxiety and Depression in Patients with Parkinson's Disease. Intern Med. 2013;52(5):539-45.
13. Skorvanek M, Gdovinova Z, Rosenberger J, Ghorbani Saeedian R, Nagyova I, Groothoff JW, et al. The associations between fatigue, apathy, and depression in Parkinson's disease. Acta Neurol Scand. 2014;131(2):80-7.
14. Braak H, Ghebremedhin E, Rüb U, Bratzke H, Del Tredici K. Stages in the development of Parkinson's disease-related pathology. Cell Tissue Res. 2004;318(1):121-34.
15. Benarroch EE, Schmeichel AM, Sandroni P, Low PA, Parisi JE. Involvement of vagal autonomic nuclei in multiple system atrophy and Lewy body disease. Neurology. 2006;66(3):378-83

16. Boeve BF, Silber MH, Saper CB, Ferman TJ, Dickson DW, Parisi JE, et al. Pathophysiology of REM sleep behaviour disorder and relevance to neurodegenerative disease. Brain. 2007;130(11):2770-88.
17. Schulz-Schaeffer WJ. The synaptic pathology of α-synuclein aggregation in dementia with Lewy bodies, Parkinson's disease and Parkinson's disease dementia. Acta Neuropathol. 2010;120(2):131-43.
18. Zhou J, Zhang J, Lam SP, Chan JW, Mok V, Chan A, et al. Excessive Daytime Sleepiness Predicts Neurodegeneration in Idiopathic REM Sleep Behavior Disorder Sleep. 2017;40(5).
19. Alves G, Larsen JP, Emre M, Wentzel-Larsen T, Aarsland D. Changes in motor subtype and risk for incident dementia in Parkinson's disease. Mov Disord. 2006;21(8):1123-30.
20. Dissanayaka NN, Sellbach A, Silburn PA, O'Sullivan JD, Marsh R, Mellick GD. Factors associated with depression in Parkinson's disease. J Affect Disord. 2011;132(1-2):82-8.
21. Grimbergen YA, Langston JW, Roos RA, Bloem BR. Postural instability in Parkinson's disease: the adrenergic hypothesis and the locus coeruleus. Expert Rev Neurother. 2009;9(2):279-90.
22. Chung S, Bohnen NI, Albin RL, Frey KA, Müller ML, Chervin RD. Insomnia and Sleepiness in Parkinson Disease: Associations with Symptoms and Comorbidities. J Clin Sleep Med. 2013;09(11):1131-7.
23. Bohnen NI, Albin RL. The cholinergic system and Parkinson disease. Behav Brain Res. 2011;221(2):564-73.
24. Kelly VE, Johnson CO, McGough EL, Shumway-Cook A, Horak FB, Chung KA, et al. Association of cognitive domains with postural instability/gait disturbance in Parkinson's disease. Park Amp Relat Disord. 2015;21(7):692-7.
25. Sulzer D, Surmeier DJ. Neuronal vulnerability, pathogenesis, and Parkinson's disease. Mov Disord. 2012;28(1):41-50.
26. Idiaquez J, Benarroch EE, Rosales H, Milla P, Ríos L. Autonomic and Cognitive dysfunction in Parkinson's disease. Clin Auton Res. 2007;17(2):93-8.
27. Malek N, Lawton MA, Grosset KA, Bajaj N, Barker RA, Burn DJ, et al. Autonomic Dysfunction in Early Parkinson's Disease: Results from the United Kingdom Tracking Parkinson's Study. Mov Disord Clin Pract. 2016;4(4):509-16.
28. Melzer TR, Watts R, MacAskill MR, Pitcher TL, Livingston L, Keenan RJ, et al. Grey matter atrophy in cognitively impaired Parkinson's disease. J Neurol Neurosurg Amp Psychiatry. 2011;83(2):188-94.
29. Zhu K, van Hilten JJ, Putter H, Marinus J. Risk factors for hallucinations in Parkinson's disease: Results from a large prospective cohort study. Mov Disord. 2013;28(6):755-62.
30. Barrett MJ, Smolkin ME, Flanigan JL, Shah BB, Harrison MB, Sperling SA. Characteristics, correlates, and assessment of psychosis in Parkinson disease without dementia. Park Amp Relat Disord. 2017;43:56-60.
31. Santangelo G, Trojano L, Vitale C, Ianniciello M, Amboni M, Grossi D, et al. A neuropsychological longitudinal study in Parkinson's patients with and without hallucinations. Mov Disord. 2007;22(16):2418-25.
32. Mata IF, Leverenz JB, Weintraub D, Trojanowski JQ, Hurtig HI, Van Deerlin VM, et al. APOE,MAPT, andSNCAGenes and Cognitive Performance in Parkinson Disease. JAMA Neurol. 2014;71(11):1405-12.
33. Alzahrani H, Venneri A. Cognitive and neuroanatomical correlates of neuropsychiatric symptoms in Parkinson's disease: A systematic review. J Neurol Sci. 2015;356(1-2):32-44.
34. Weintraub D, Newberg AB, Cary MS, Siderowf AD, Moberg PJ, Kleiner-Fisman G, et al. Striatal dopamine transporter imaging correlates with anxiety and depression symptoms in Parkinson's disease. J Nucl Med. 2005;46(2):227-32.
35. Bohnen NI, Kaufer DI, Hendrickson R, Constantine GM, Mathis CA, Moore RY. Cortical cholinergic denervation is associated with depressive symptoms in Parkinson's disease and parkinsonian dementia. J Neurol Neurosurg Amp Psychiatry. 2007;78(6):641-3.
36. Kang JH, Mollenhauer B, Coffey CS, Toledo JB, Weintraub D, Galasko DR, et al. CSF biomarkers associated with disease heterogeneity in early Parkinson's disease: the Parkinson's Progression Markers Initiative study. Acta Neuropathol. 2016;131(6):935-49.
37. Ramakers IH, Verhey FR, Scheltens P, Hampel H, Soininen H, Aalten P, et al. Anxiety is related to Alzheimer cerebrospinal fluid markers in subjects with mild cognitive impairment. Psychol Med 2012 Sep;43(5):911-20.
38. Doty RL. Olfaction in Parkinson's disease and related disorders. Neurobiol Dis. 2012;46(3):527-52.
39. Tessa C, Lucetti C, Giannelli M, Diciotti S, Poletti M, Danti S, et al. Progression of brain atrophy in the early stages of Parkinson's disease: A longitudinal tensor-based morphometry study in de novo patients without cognitive impairment. Hum Brain Mapp. 2014;35(8):3932-44.
40. Jia X, Liang P, Li Y, Shi L, Wang D, Li K. Longitudinal Study of Gray Matter Changes in Parkinson Disease. Am J Neuroradiol. 2015;36(12):2219-26.