Data-Driven Discovery of Movement-Linked Heterogeneity in Neurodegenerative Diseases

Mark Endo[1], Favour Nerrise[2], Qingyu Zhao[3], Edith V. Sullivan[4],
Li Fei-Fei[1], Victor W. Henderson[5,6], Kilian M. Pohl[2,4],
Kathleen L. Poston[5], Ehsan Adeli[1,4*]

[1]Department of Computer Science, Stanford University, Stanford, CA, USA
[2]Department of Electrical Engineering, Stanford University, Stanford, CA, USA
[3]Department of Radiology, Weill Cornell Medicine, New York, NY, USA
[4]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA
[5]Department of Neurology and Neurological Sciences, Stanford University, Stanford, CA, USA
[6]Department of Epidemiology and Population Health, Stanford University, Stanford, CA, USA

Word Count: 3998

Abstract

Neurodegenerative diseases manifest different motor and cognitive signs and symptoms that are highly heterogeneous. Parsing these heterogeneities may lead to an improved understanding of underlying disease mechanisms; however current methods are dependent on clinical assessments and somewhat arbitrary choice of behavioral tests. Herein, we present a data-driven subtyping approach using video-captured human *motion* and *brain* functional connectivity (FC) from resting-state (rs)-fMRI. We applied our framework to a cohort of individuals at different stages of Parkinson's disease (PD). The process mapped the data to low-dimensional measures by projecting them onto a canonical correlation space that identified three PD subtypes: Subtype I was characterized by motor difficulties and poor visuospatial abilities; Subtype II exhibited difficulties in non-motor components of activities of daily living and motor complications (dyskinesias and motor fluctuations); and Subtype III was characterized by predominant tremor symptoms. We conducted a convergent validity analysis by comparing our approach

*Correspnoding Author: eadeli@stanford.edu

to existing and widely used approaches. The compared approaches yielded subtypes that were adequately well-clustered in the motion-brain representation space we created to delineate subtypes. Our data-driven approach, contrary to other forms of subtyping, derived biomarkers predictive of motion impairment and subtype memberships that were captured objectively by digital videos.

Keywords: Parkinson's disease, subtyping, functional connectivity, human motion analysis, canonical correlation analysis, digital biomarkers

1 Main

Neurodegenerative diseases, including Parkinson's disease (PD), are highly heterogeneous in their clinical manifestations. Deciphering the variable expression in individuals may reveal clues to the underlying disease mechanisms [17, 41]. In this work, we aimed to identify sources of this heterogeneity from a data-driven perspective by directly studying movement from videos and brain function from resting-state functional MRI (rs-fMRI) analysis. Accordingly, we developed a subtyping approach that groups individuals based on raw motion and brain functional connectivity (FC) data. Joint analysis of these two modalities of data offers a comprehensive perspective on the intricate link between brain function and movement impairments [56, 49, 10]. Leveraging the recent advances in Artificial Intelligence (AI) for encoding 3D human motion from videos [15, 30] and brain connectivity patterns from rs-fMRI [38], we correlated motor-brain encodings to develop precise subtypes. Such a subtyping scheme could enhance diagnostic accuracy, elucidate diverse disease pathologies, foster the development of tailored therapeutic interventions, and engender effective and individualized patient care. To demonstrate the capability of the methods, we apply them to detect PD subtypes and individualized biomarkers.

Previous work on data-driven PD subtyping typically used cluster analysis derived from clinical variables, such as motor and non-motor symptoms and signs observed by neurologists, to determine which combination of variables best differentiated the subtypes with the ultimate goal of evaluating whether the clusters were meaningful and clinically interpretable [53]. Clinical variable clustering, however, is known to be highly subjective [32, 14] even when this approach has moved past conventional, rule-based methods. A common rule-based example divides people into tremor-dominant (TD) or postural instability and gait difficulty (PIGD) subtypes based on the ratio of the mean Movement Disorders Society-Unified Parkinson's Disease Rating Scale (MDS-UPDRS) tremor scores to the mean MDS-UPDRS PIGD scores (noted as TD/PIGD) [50, 54]. Recent works have included objective measurements of data for subtyping such as biospecimen examinations, neuroimaging measures, genomic data, neurophysiological assessments, and REM sleep behavior [60, 3, 7]. By contrast, our proposed approach moved entirely away from clustering based on clinical variables and instead used objective, raw motion and FC data (Figure 1(A)). Doing so enabled us (1) to build a latent space that correlated

brain function with motion and mobility, thereby facilitating the identification of disease mechanisms, and (2) to extract markers from gait videos and rs-fMRIs that correlated with disease severity and our generated subtype assignments. In addition to the identified brain connectivity markers, the features extracted from digital videos could serve as potential digital biomarkers, and their trajectory of changes could be tracked with treatment or disease progression.

Our data-driven approach explored relations between motion features and FC subnetworks that highlighted specific brain network disruptions associated with movement-linked impairments. When applied to a cohort of individuals with PD at different levels of movement-impairment severity, our method resulted in three clinical subtypes each characterizing different aspects of the disease such as motor experiences of daily living, visuoperceptual abilities, tremor, non-motor aspects of daily living, and motor complications of dopaminergic therapy (Figure 1(C)). In addition, we demonstrated convergent validity by correlating our subtyping approach with existing subtyping approaches. Lastly, we derived digital biomarkers for PD motor impairment and subtype membership based on motion and FC data.

2 Results

2.1 Motion features predictive of motor impairment severity

We used a Transformer-based model [15] pre-trained on motion forecasting to predict motor impairment severity from videos of gait examinations, following our prior work [15]. This pre-trained method accurately predicted gait impairment severity with an F_1 score of 0.76, precision of 0.79, and recall of 0.75 (performance of 0.25 implies a random classifier). Its average AUC across the gait severity labels (one-vs-rest) was 0.80 (see Table 1 and Supplementary Table 1. The benefits of this approach were (1) it learned useful motion representations using motion forecasting as a self-supervised task leveraging public motion capture datasets, and (2) it enabled joint training of the motion forecasting and gait impairment score prediction branches that ultimately improved the extraction of meaningful motion representations (i.e., the Motion Encoder in Figure 1). In the subsequent sections, we combined these representations with FC features and built a low-dimensional latent space, in which these combined representations were highly correlated.

2.2 Functional connectivity predicting motor impairment severity

We experimented with two approaches for choosing salient FC subnetworks that we combined with motion features to generate subtype representations. The first approach used connections identified as strongly correlated (two-sided p-value < 0.001 from Pearson correlation coefficient) with motor impairment (MDS-UPDRS Section 3.10) in our dataset. The resulting five connections, listed in descending

order based on their correlation with motor impairment, are as follows: Cerebellum 7b Left ↔ Lateral Occipital Cortex inferior division Left, Cerebellum Crus1 Left ↔ Frontal Orbital Cortex Left, Cerebellum 3 Right ↔ Heschl's Gyrus Right, Cerebellum 6 Left ↔ Frontal Orbital Cortex Left, and Default Mode Network LP Left ↔ Temporal Fusiform Cortex anterior division Right. In addition to the selected connections, we used all FCs from across the whole brain and trained a simple neural Network (i.e., MLP) and a graph convolutional neural network (GCN) [29] to predict the gait impairment scores (see Table 1). Compared to the other complicated methods, a simple model using these five connections as input predicted motor impairment significantly better than the null hypothesis based on prediction using a random classifier (in leave-one-out cross-validation settings).

The second approach only considered connections from subnetworks that were previously found to have effects on motor outcomes based on the literature: the cerebellar-prefrontal network (connections with motor planning and control) [35, 33, 18, 62], cerebellar-motor network (connections with gait speed, balance, turning, and torso inclination) [2, 22, 39], and pallidal-sensorimotor network (connections with dynamic gait features and sensorimotor integration) [28] (see Supplementary Information for more information about the selection of these networks). The inclusion of this smaller set of connectivities reduced the number of correlations to test. To maintain consistency with the first approach which selected five connections, we chose the top five correlations from the three subnetworks to use as the FC measures in the second approach. These five connections most correlated with motor impairment were all from the cerebellar-prefrontal subnetwork. These top five connections were Cerebellum Crus1 Left ↔ Frontal Orbital Cortex Left, Cerebellum 6 Left ↔ Frontal Orbital Cortex Left, Cerebellum Crus1 Left ↔ Frontal Orbital Cortex Right, Cerebellar Posterior ↔ Frontal Orbital Cortex Left, and Cerebellum 6 Right ↔ Frontal Orbital Cortex Left. These top five connections for the second approach highlight motion-linked impairments specific to the cerebellar-prefrontal network.

In addition, for each subnetwork, we evaluated how well a model could predict motor impairment. Of the three subnetworks tested, the cerebellar-motor and cerebellar-prefrontal networks predicted the motor impairment scores better than the setup where connections were randomly selected from across the whole brain (p-value < 0.05 using Wilcoxon signed-rank test). Use of the five cerebellar-prefrontal network connections most strongly correlated with motor impairment as input into the model resulted in improved prediction accuracy compared to the use of the entire subnetwork. Results are summarized in Table 1 (see Supplementary Table 1 for detailed results).

2.3 Motion-FC subtypes

After projecting the motion representations and FC subnetworks onto a canonical correlation space and subtyping individuals based on their locations in this space (shown in Figure 1(B)), we examined the characteristics of the resulting subtypes. We characterized the subtypes by identifying variables that

exhibited significantly different expression levels across subtypes (p-value < 0.05 from Chi-square test for categorical variables and from the Kruskal-Wallis test for continuous variables). The tested variables included demographics (e.g., sex and age), MDS-UPDRS scores, cognitive tests (e.g., Montreal Cognitive Assessment, MoCA, [37] and Judgment of Line Orientation, JLO, [5]), and extracted measurements from gait videos, including walking time, walking speed, and average torso angle.

Subtype I was characterized by difficulties in motor abilities in daily living (including eating, personal hygiene, turning over in bed, walking/balance, and freezing from MDS-UPDRS Part-II), bradykinesia/rigidity, and gait difficulties on the exam (including pronation-supination of hands, arising from a chair, postural instability, and poor posture from MDS-UPDRS Part-III), and lower visuoperceptual abilities as measured by the Judgment of Line Orientation (JLO) test [5]. Subtype II was marked by difficulties in non-motor experiences of daily living (including increased depression and fatigue from MDS-UPDRS Part-I) and more motor complications of dopaminergic therapy (including time spent with dyskinesias, functional impact of dyskinesias, and functional impact of fluctuations between ON and OFF medication states from MDS-UPDRS Part-IV). Subtype III exhibited predominant tremor symptoms (including amplitude of rest tremor in upper and lower extremities and consistency of rest tremor from MDS-UPDRS Part-III). Not surprisingly, there was also a significant difference in age across the three subtypes, given that older age is a common factor in declining mobility [42]. The traditional clinical variable-based subtyping approach also differed significantly in age across subtypes, as summarized in Supplementary Figure 2. Previous work also found that age and onset age differ across subtypes, as onset age is a traditional way of subtyping PD [60, 44]. A heatmap visualization comparing mean values of the significant variables across subtypes is presented in Figure 1(C), and a full comparison of distributions of significant variables across subtypes is displayed in Extended Data Figure 1.

2.4 Convergent validity and subtype approach comparison

We assessed the convergent validity of our subtyping approach by measuring the correlation between our subtyping and longstanding subtyping approaches using the metric σ, which measures within-cluster variance in the representation space (see Figure 2(A)) created by our method. One of these approaches computes cluster analysis on a broad selection of clinical variables, while the other approach uses tremor and gait scores from MDS-UPDRS to classify individuals into subtypes. This comparison confirms that the representation space (and hence the clusters) generated by our data-driven method align with those produced by traditional subtyping approaches, demonstrating convergent validity. We further qualitatively compared the subtype characteristics of our approach to the characteristics of the clinical variable-based approach, outlined in the following subsections.

2.4.1 Clinical variable-based subtyping

We first compared our approach to the common subtyping approach of conducting cluster analysis on a set of clinical variables. We measured the correlation between the two approaches using the metric σ and then compared this score to the correlation between our approach and random subtype groupings. The conventional subtypes had a higher clustering in our representation space than 97.6% of random individual groupings. This high correlation demonstrates that our generated representation space captured information about individuals that the conventional clinical variable-based subtyping approach used to subtype. It also signifies that the conventional subtypes are meaningfully separated in our representation space. Figure 2(A) includes a histogram of random group clustering scores, illustrating that conventional subtypes are more highly clustered than the majority of randomly assembled groupings. The graphical visualization showing how well clinical variable-based subtypes are grouped in our representation space is shown in Figure 2(B) top-row.

In addition, we qualitatively analyzed the similarities between the characteristics of this clinical variable-based approach and our approach and found that the two approaches resulted in similar subtype profiles. For example, Subtypes I in both approaches were described by motor impairment in activities of daily living and poor motor exam. Additionally, Subtype III in both approaches exhibited overall mild bradykinesia/rigidity and gait impairment but more rest tremor consistency. However, there were differences between the characteristics of the two approaches. For example, Subtype II in the conventional approach was characterized by low cognitive function as measured by the MoCA [37], California Verbal Learning Test (CVLT) [12], and Symbol Digit Modalities Test (SDMT) [46]. In our approach, Subtype I exhibited cognitive impairment as measured by JLO [5]. Generally, the clinical variable-based approach contained many variables relating to cognition, whereas our data-driven approach contained only one. Full subtype profiles for the clinical variable-based approach are in Supplementary Figure 2.

2.4.2 Conventional TD/PIGD subtyping

We additionally compared our approach with the conventional tremor-dominant/postural instability and gait difficulty (TD/PIGD) subtyping approach. Using the same setup as the clinical variable-based subtyping comparison, we found that TD/PIGD subtypes had a better clustering score than 99.9% of random individual groupings. This outcome again supports the conclusion that the representation space created by our entirely data-driven approach exhibits an interpretable space. As another measure of convergence, we found that the TD/PIGD ratio variable exhibited significantly different expression levels across our subtypes (p-value < 0.05). Lastly, we visualized the clustering of TD/PIGD subtypes in our subtyping space of motion features in Figure 2(B) bottom-row. We found that the PIGD and TD individuals were clustered on different sides of the representation space. As expected, the indeterminate subtype was clustered in between the PIGD and TD subtypes.

2.5 Subtype (digital) biomarker discovery

Next, we analyzed how masking individual joints by exclusion and removing functional connectivities affected 1) the ability of models to estimate gait impairment severity and 2) subtype generation as measured by subtype grouping similarity. This procedure identified the 'impact' or the importance of each body joint or FC measure. For predicting gait impairment, masking the elbows had the greatest negative effect on predictive performance, with masking of wrists and ankles also reducing predictive performance. For subtyping, we found that movement of wrists, knees, and the spine had the most effect on grouping individuals. On the FC side, out of the five connections used for subtyping, removing the Default Mode Network-Fusiform Cortex connection had the greatest impact on final groupings. Figure 3 contains visualizations of the statistical relevance of each joint and each FC effect on subtype grouping.

We also calculated various metrics such as movement variability, velocity, and axis angle on the joints that had the most effect on estimating gait impairment severity. We found that individuals with gait impairment had significantly different expressions of selective motion variables than individuals without gait impairment. For example, individuals with gait impairment had lower elbow rotational velocity, lower wrist velocity, and lower ankle rotational velocity, but higher ankle z-axis movement variability. The complete set of results is shown in Figure 3 with distributions displayed in Extended Data Figure 2.

3 Discussion

We designed an objective, data-driven subtyping approach derived from videos of individuals performing clinical tests and from resting-state functional MRIs to identify factors that contribute to movement-linked heterogeneity of neurodegenerative diseases. Instead of subtyping based on a subjective selection of variables recorded by clinicians, our work was grounded in inherent differences in PD expression acquired through features from motion and FC data. Correlations between these two sets of features were used to generate a readily interpretable representation space (Figure 2(B)) for subtyping. Thus, the subtypes were built upon motor-brain links discovered in the data without requiring clinician input. We hypothesize that these motor-brain links account for substantial variance in defining PD heterogeneity. Our data-driven approach resulted in newly identified digitally trackable biomechanical markers that can be continuously monitored (Figure 3(B)) and deployed in a mobile app. With continuous monitoring of gait markers, patients could potentially receive individualized feedback directed by changes in their motor-brain links [20].

When applied to a cohort with clinical PD diagnosis, our modeling identified novel PD subtypes with distinct characteristics completely in a data-driven fashion (Figure 1(C)). Subtype I was characterized by difficulties in motor activities of daily living, poor motor test performance, and compromised vi-

suoperceptual abilities; Subtype II exhibited difficulties in non-motor aspects of daily living and salient motor complications of dopaminergic therapy; and Subtype III was characterized by tremor but otherwise lowest symptom severity. Age was also a significant spectrum characteristic, where Subtype I was an older cohort and Subtype II was a younger cohort. A similar pattern held with onset age where the first subtype had an older onset age, the second subtype had a younger onset age, and the third subtype had an onset age in between the first and second subtypes (see Extended Data Figure 3). Using onset age is another traditional way of PD subtyping [44]. In line with previous literature, we observed that the second subtype (with the youngest onset age) exhibited the most depression and had a high frequency of treatment-related dyskinesia [34]. Also consistent with prior studies, our Subtype I had the latest age onset and presented a more severe motor phenotype than other subtypes [40]. Although previous work reported that tremors become more common with older onset age, our Subtype III, which had age onset between the other two subtypes, was the group with the greatest tremor symptoms [58].

Age was not a strong correlate of disease severity, as Subtype III had the mildest observed symptom severity despite being older than Subtype II. We observed that the youngest subtype had more motor complications of dopaminergic therapy, consistent with prior studies [55]. Critically, disease duration was not a defining characteristic of our subtyping approach (p-value = 0.112), demonstrating that our work was revealing heterogeneous subtypes and not solely different disease stages. Disease duration only became significant in its interaction with age through onset age. The motor impairment severity score measured clinically with MDS-UPDRS Section 3.10 that we used to train the motion model and choose FC subnetworks was also not a defining characteristic (p-value = 0.423).

We demonstrated the convergent validity of our approach by measuring its correlation with existing subtyping approaches. Both clinical variable-based and TD/PIGD subtyping approaches exhibited a high correlation with our approach as measured by their clustering scores in our representation space (Figure 2(A)). This high clustering demonstrated that our low-dimensional representation space captured clinically relevant components of PD heterogeneity. Despite the high clustering of traditional subtyping approaches in our representation space, our approach used different criteria from other approaches. As such, our subtyping yielded groupings of individuals that were notably distinct from those of the conventional subtyping approaches. As seen in Figure 2(B,C), subtypes of our approach were partitioned into the bottom left, bottom right, and top regions of the subtyping space whereas the clinical variable-based subtypes were roughly partitioned into the bottom, middle at top regions and the TD/PIGD subtypes were partitioned into the bottom and top regions.

Qualitatively comparing the subtype characteristics of our approach and the clinical variable-based approach revealed that the two approaches shared many characteristics. For example, the first subtype in both approaches exhibited more motor difficulties, and the third subtype displayed overall less severe symptom expression. However, the clinical variable-based approach had more differentiated variables,

particularly related to cognition. Our approach based on video motion data and brain functional patterns only resulted in one cognitive test (JLO [5]) with significantly different scores across subtypes, whereas the clinical variable-based approach had 11 differentiated cognitive tests (MoCA, BVMTR, CVLT, JLO, HVOT, SDMT, FAS, Animals, Trails, SCWT, and BNT [37, 4, 12, 5, 23, 46, 47, 43, 6, 48, 26]). Although the clinical variable-based approach captured cognitive impairments to a high degree, it did so by directly using these cognitive tests to subtype patients, thereby using the cognitive grouping as both a dependent and independent variable. Overall, our findings demonstrate the effectiveness of the data-driven approach combining learned motion features from gait examination videos with function connectivities for discovering motor-brain links independent of predetermined groupings or comprehensive clinical examinations. Thus, our approach provides a valuable method for objectively exploring PD heterogeneity, which could be simply extended to the analysis of different neurodegenerative diseases. Whether the constellations of cognitive test sensitivity change with advancing disease or evolving dementia disorders remains to be tested with longitudinal data.

Our data-driven approach enabled the discovery of digital *mechanobiomarkers* relating to joint movement and imaging *neurobiomarkers* associated with functional connectivities (Figure 3):

- **Motion**: Our method found that elbows, wrists, and ankles had a high impact on predicting gait impairment, whereas elbows, knees, and the spine had the most effect on ultimate subtyping assignments. In addition to examining the overall effect of joints on motor impairment protection and subtype assignment, we tested the hypothesis that models learn patterns of joint movement and rotation to discern PD gait impairment severity (Extended Data Figure 2. Supporting this hypothesis, individuals with gait impairment had lower elbow and ankle rotation velocity, which comports with previous research showing that PD reduces the total excursion of elbow and ankle motion [36]. These walking patterns may be critical for recognizing gait impairment, and masking these joints attenuated the model's predictive abilities. Note that our motion model learned from automatically generated skeleton estimations of individuals extracted from raw videos instead of precise motion capture data, which makes the method plausible for use in simple home or clinical settings. Further, our analysis found that individuals with gait impairment had higher foot lift (ankle z-axis movement variability), but previous research noted that individuals with PD have lower heel elevation [36]. One possibility is that the high footlift we observed in individuals with gait impairment was due to less smoothness (greater jittering) of recovered poses for those individuals [30].
- **FC**: As illustrated in Section 2.5 and Figure 3(B), our approach identified the FC between the Default Mode Network (DMN) Lateral Parietal (LP) Cortex Left and the Temporal Fusiform Cortex anterior division Right as the most significant contributor to our subtyping scheme. While the correlation between DMN disconnection and cognitive dysfunction in individuals with PD is

widely established [13], the full extent and nature of DMN function in PD, especially in aspects beyond cognitive function, remains unclear [61]. As PD individuals with primary akinetic/rigidity are likely to develop cognitive deficits, previous works have examined changes in the DMN between different motor subtypes [27]. For example, a study found that compared to healthy controls, PD individuals with akinetic rigidity that are still cognitively unimpaired exhibited prominently decreased connectivity between the posterior DMN (encompassing the lateral parietal cortex) and the fusiform gyrus [24]. Our data-driven approach identified this same connectivity constellation as a distinguishing element among our subtypes. Given our generated subtypes were characterized by numerous motor differences, our finding that the DMN-Fusiform Cortex connection had the greatest impact on subtype groupings is intriguing. This is validated by the literature that the fusiform cortical role in processing complex visual information can be directly related to mobility and gait, as effective navigation through an environment requires recognizing and interpreting various visual cues. Additionally, the DMN, while typically less active during physical tasks, interacts with networks that control attention and the executive function [8, 21]. These functions are essential for planning movements, making decisions about where to go, and how to navigate obstacles. Disruptions in the DMN have been associated with difficulties in attention and memory, which can indirectly affect mobility and gait [11]. More targeted research is required to further unravel the complexities of DMN dysfunction in PD, particularly relating to tremors, gait difficulties, and psychiatric disturbances. The other four connections used for subtyping all involved the cerebellum. Previous work has found PD-related pathological changes in the cerebellum, and these changes related to akinesia/rigidity, tremor, gait disturbance, dyskinesia, and some non-motor symptoms [59]. Complementary to these findings, our data-driven analysis identified the cerebellar-occipital connection as the connection with the highest correlation with gait impairment severity.

One limitation of our work was the reliance on collecting gait impairment severity scores for training the motion encoder and choosing salient FC subnetworks. Presumably, using this gait impairment severity score to train our motion encoder and dictate FC subnetwork selection also increased subjectivity as the score is based on one aspect of PD expression and is recorded by a clinician. Another limitation was that our approach used fMRI scans for subtyping individuals, which required specific machinery and trained medical staff for data collection. This study further requires access to longitudinal data to track the change of subtype assignments, treatment utility, or disease progression over time, which defines directions for future work. In addition, since we employed modality-specific feature extractors for motion and brain FC data and later projected them in a shared space, we could not explore deeper connections between motion and fMRI biomarkers. Using data-driven approaches such as AI algorithms on smaller-scale datasets, particularly in the setting of neurodegenerative disease, presents challenges such as the potential to overfit. To address these challenges, in this work, we (1) first pre-trained our

motion encoder on a large-scale public 3D motion capture dataset to learn useful motion representations, (2) followed a leave-one-out cross-validation setup, and (3) chose a simple method for the modeling of functional connectivity. As a result, this work demonstrates how to make use of large-scale public motion capture datasets to translate the knowledge for modeling motion for clinical applications.

In conclusion, we subtyped individuals with PD according to their representations based on motion and FC data. With our data-driven approach, the resulting subtypes defined characteristics that were readily interpretable and exhibited convergent validity without relying on extensive clinical testing. Finally, use of raw data led to the discovery of digital mechanobiomarkers defining PD gait impairment severity and objective PD subtyping.

4 Methods

4.1 Participants and dataset

We acquired 31 samples of rs-fMRIs, video recordings of MDS-UPRDRS part 3 exams, MDS-UPDRS part 1-4 scores, and cognitive test scores from participants with a PD diagnosis. Specifically for section 3.10 of the MDS-UPDRS exam (gait impairment severity), we had three board-certified movement disorders neurologists score each video recording (details explained in [32]). In this section of the exam, participants were recorded from the frontal perspective. They were instructed to walk at least 10 meters towards the camera, then turn around and walk back. According to the MDS-UPDRS criteria, the neurologists scored gait impairment based on many behaviors including stride amplitude, stride speed, the height of foot lift, heel strike during walking, turning, and arm swing [19]. We additionally obtained gait video recordings from several control non-PD subjects, of which we selected 23 who were frequently matched by age and sex with the PD cohort. For Sections 2.3, 2.4, and 2.5, we excluded one example as it was missing MDS-UPDRS part 1-4 scores. All data collection procedures were approved by the Stanford IRB and all individuals consented to participating in the study. In collecting, processing, and analyzing the data, we complied with all relevant ethical regulations. Score distribution and other information about the study participants are included in Table 2 and Extended Data Figure 4. To pre-train our motion encoder, we used NTU RGB+D dataset [45], which contains 3D skeletal data for 56,880 video samples.

4.1.1 Data preprocessing

Videos. The sections of the videos documenting gait examinations were selected where participants are instructed to walk directly toward and away from the camera. Due to the difficulty of obtaining video recordings for participants with extreme motor impairment, we combined gait impairment scores 3 and 4. In order to extract 3D skeletons from videos of gait examinations, we utilized Video Inference

for Body Pose and Shape Estimation (VIBE) [30]. This method estimates 3D human meshes from videos and regresses them to skeletons with 49 predefined joints. We utilized 17 of these joints which correspond to NTU joints used to pre-train the Transformer-based motion encoder [15]. Note that these extracted 3D skeletons from videos are subject to estimation errors compared to approaches such as multi-camera setups or motion capture systems based on Inertial Measurement Units. However, our 3D estimation algorithms exhibited reliable results after manual inspection of the reconstructed 3D meshes, providing a contactless and affordable approach for recording movement in everyday clinics. From the 3D human meshes, we additionally extracted rotational angles in the form of relative rotation of joints in axis-angle format which were used in Section 2.5.

rs-FMRIs. For each individual, pre-processed functional connectivity (FC) matrices were obtained using a combined Harvard-Oxford and AAL parcellation atlas [57] with 165 regions of interest (ROIs), where each the (i, j)th entry in the matrix is the Pearson correlation between the average rsfMRI signal measured in ROI i and ROI j. The FC matrices were corrected for possible motion artifacts using the CONN toolbox [57].

4.2 Data-Driven Motion-FC PD subtyping

4.2.1 Motion encoder

The motion representations reduced the videos of individual's gait examinations to a vector using a deep neural network trained on various motion prediction tasks. We used intermediate model representations for subtyping. Specifically, we began by training a model on the task of human motion forecasting using large-scale public 3D motion capture data. In this task, representations of past movements were used to forecast future movements. Specifically, given a sequence of t 3D skeletons $\mathbf{x}_{1:t}$, a model predicted the next M skeletons $\mathbf{x}_{t+1:T}$. We used recent advances in computer vision and developed an encoder-decoder Transformer model. We had previously shown that the model could learn useful representations from motion forecasting of public motion dataset that can be applied to the task of gait impairment severity estimation [15]. In short, an encoder took as input a sequence of skeleton representations and computed intermediate, latent representations. Then, the intermediate representations were fed into a decoder to output future skeleton predictions.

We then fine-tuned the model on videos of gait examinations, additionally predicting the gait severity impairment score by feeding the intermediate representations through a simple linear classifier. Note that fine-tuning refers to the process of adapting the trained model to our application, translating the technology to the specific clinical application of interest.

4.2.2 FC encoder

The FC component also comprised an encoder that reduced the fMRIs to a vector by selecting subnetworks that are highly correlated with motor symptoms of PD (within cross-validation settings). We followed two different approaches for selecting relevant subnetworks for subtyping. In the first approach, we determined how correlated each connectivity in the rs-fMRI data was with gait impairment severity by calculating Pearson's correlation coefficients. We then selected connections for subtyping that had an associated p-value of less than 0.001. In the second approach, we took note of previous literature that discusses how certain subnetworks of the brain have effects on various motor outcomes and only considered connections in particular subnetworks. Based on previous literature, we looked at connections from the cerebellar-prefrontal network, cerebellar-motor network, and pallidal-sensorimotor network. To analyze the relevance of each subnetwork for motor impairment, we trained a model for each subnetwork that had subnetwork connections as input and predicted gait impairment severity using a simple Multi-Layer Perceptron (MLP). We used a leave-one-out cross-validation setup and trained each fold for 100 epochs. For the whole-brain setup, we additionally used a Graph Convolutional Network (GCN) [29]. Looking into more detail about individual connections instead of entire subnetworks, we also analyzed how highly correlated individual connections are with motor impairment. We found that out of the top 5 connections most correlated, all were from the cerebellar-prefrontal network. We used these top 5 connections as brain features in the second approach.

To further demonstrate that FC data contains detectable patterns that can be learned to accurately predict PD motion impairment, we utilized our recent geometric attention-based model (xGW-GAT) [38] for gait impairment severity estimation. This method represented brain connectivities as a learnable graph structure to characterize discriminative attributes of edge encodings for gait impairment severity estimation. The method used a stratified, learning-based sample selection method to mitigate the small dataset size and label imbalance. Note that this method followed a transductive learning [25] approach, so intermediate representations were not generalizable to other tasks such as subtyping. Therefore, we used the simpler and more interpretable approach of selecting FC subnetworks as our representations used for subtyping.

4.3 Subtype clustering

Given motion and brain representations, we created individual representations by using canonical correlation analysis (CCA) [51] to get low-dimensional, correlated motion and FC features. In our work, we used two components from motion and FC features, adding up to a combined representation space of 4 dimensions. With these features, we then generated subtypes by grouping CCA representations using k-means clustering. We utilized the standard Lloyd's algorithm version of k-means clustering [31]

with initial clusters chosen using k-means++ [1]. In the assignment step, each individual was assigned to the cluster with the closest centroid (least squared Euclidean distance). Mathematically, we denote cluster assignment for cluster C_i as

$$C_i = \{l : \|\mathbf{x}_l - \mu_i\|^2 \leq \|\mathbf{x}_l - \mu_j\|^2 \forall j, 1 \leq j \leq k\},$$

where each l is an individual, $\mathbf{x}$ is a matrix representing the CCA features for all individuals, μ is a matrix representing cluster centroids, and $\|\cdot\|$ is the ℓ_2 norm. In the update step, the centroids were recalculated for each cluster C_i as $\mu_i = \frac{1}{|C_i|}\sum_{l \in C_i} \mathbf{x}_l$. This technique attempted to minimize the within-cluster sum of squares (variance), which is denoted as

$$\sigma(C) = \sum_{i=1}^{k} \sum_{l \in C_i} \|\mathbf{x}_l - \mu_i\|^2.$$

For generating subtypes, we calculated the optimal number of clusters using Hartigan's rule with a threshold of 12 following [9] (see Supplementary Figure 7 for a visualization of σ scores and Hartigan's statistic values as the number of subtypes increases). Along with using σ to generate subtypes, we also used σ as a metric for how well any given subtype grouping D was clustering according to our generated motion/fMRI representation space, represented as $\sigma(D)$.

For calculating which clinical variables were differently expressed among subtypes, we used chi-square tests for categorical variables and Kruskal-Wallis tests for non-normal continuous variables, following [60].

4.4 Motion-FC subtype characteristics

For determining subtype characteristics, we identified variables that exhibited significantly different expression levels (p-value < 0.05) across subtypes. For categorical variables (e.g., medication state, dyskinesia) and ordinal variables with no more than five categories (e.g., gait, posture), we used the Chi-square statistical test. Meanwhile, for continuous variables (e.g., age, PIGD/TD) and ordinal variables with more than five categories (e.g., MoCA, MDS-UPDRS Part-I), we used the Kruskal-Wallis statistical test. For variables that exhibited significantly different expression levels according to these tests, we also displayed subtype z-scores in Figure 1(C). A subtype z-score for subtype i is defined as $\frac{\mu_i - \mu}{\sigma}$, where μ_i is the variable mean value for the subtype, μ is the variable mean value across all subtypes, and σ is the variable standard deviation across all subtypes.

4.5 Subtype (digital) biomarker discovery

To analyze the importance of each joint in model prediction and subtype generation, for each joint, we set its value to 0 across all frames and trained the deep learning model to predict motor impairment. We followed a similar approach for FC data by removing each connection. We measured the effect of each joint/connection on the model's predictive ability by subtracting the model's F_1 score with the masked joint/connection from the model's F_1 score using all data. For the subsequent subtype generation, we used a clustering similarity metric [52] to determine how much masking joints/connections affect subtype clustering. Specifically, we define subtype groups from our approach as $C = \{C_1, C_2, ..., C_m\}$ and the subtype groups from the modified approach which masks a joint/connection as $D = \{D_1, D_2, ..., D_n\}$. We then define the similarity matrix for C and D as

$$S_{C,D} = \begin{bmatrix} S_{11} & S_{12} & ... & S_{1n} \\ S_{21} & S_{22} & ... & S_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ S_{m1} & S_{m2} & ... & S_{mn} \end{bmatrix}$$

where S_{ij} is Jaccard's Similarity Coefficient or p/q where p is the size of intersection and q is the size of union of the sets C_i and D_j. The similarity of groupings C and D is then defined as $\text{Sim}(C, D) = \sum_{i \leq m, j \leq n} S_{ij} / \max(m, n)$. We calculated the effect of joint/connection masking as subgroup dissimilarity or $1 - \text{Sim}(C, D)$.

Data availability

We provided a toy dataset of select motion encoding outputs and fMRI subnetworks with our code release [16]. To protect study participant privacy, we are unable to release the clinical data or the gait examination videos. The NTU RGB+D dataset [45] used to pre-train the motion encoder model is available at `https://rose1.ntu.edu.sg/dataset/actionRecognition/`.

Code availability

We have released code for subtype analysis and predicting motor impairment using FC data [16]. The previously published code for the GaitForeMer motion encoder is available at `https://github.com/markendo/GaitForeMer`.

Acknowledgements

This research was supported in part by NIH grants AA010723 (E.V.S.), AA017347 (E.V.S., E.A.), AG047366 (V.W.H., K.L.P., E.A.), MH113406 (K.M.P.), and AG066515 (V.W.H.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This study was also supported by the Stanford School of Medicine Department of Psychiatry & Behavioral Sciences Jaswa Innovator Award (E.A.) and the Stanford Institute for Human-centered Artificial Intelligence (HAI) GCP Cloud Credit (E.A.).

Author Contributions Statement

M.E.: Methodology, Investigation, Software, Visualization, Writing - original draft; F.N.: Methodology; Q.Z.: Methodology, Writing - review & editing; E.V.S.: Methodology, Writing - review & editing; L.F.: Methodology; V.W.H.: Conceptualization; K.M.P.: Methodology, Writing - review & editing; K.L.P.: Conceptualization, Methodology, Writing - review & editing; E.A.: Conceptualization, Project administration, Supervision, Writing - original draft, review & editing.

Competing Interests Statement

The author(s) declare that they have no competing interests.

Tables

Table 1: Results on how accurately human motion (video) or FC (rs-fMRI) data can be used to predict motor impairment. We followed a leave-one-out cross-validation approach. Note that the reported video motion analysis results additionally include control data. For FC data, we used a simple MLP for predictions and averaged results across 20 runs. The 100 Random connections are chosen from the whole brain 20 times, creating a null hypothesis as a baseline for comparisons. Out of the three tested subnetworks, the cerebellar-prefrontal network exhibited the highest predictive ability of motor impairment. Going beyond connections from the cerebellar-prefrontal subnetwork and using connections from across the whole brain, with both MLP and GCN [29], yielded similar performance. * indicates significantly better than chance (p-value < 0.05 using one-sided Wilcoxon signed-rank test). The p-values for the cerebellar-motor, cerebellar-prefrontal, top 5 cerebellar-prefrontal, and top 5 whole-brain setups were 0.029, 9.54×10^{-7}, 9.54×10^{-7}, and 9.54×10^{-7}, respectively.

Approach	Average AUC (one-vs-rest)
Video motion analysis*	0.804
Functional networks (× 20 Runs)	
Pallidal-sensorimotor (2×3 connections)	0.427
Cerebellar-motor (20×3 connections)*	0.519
Cerebellar-prefrontal (20×5 connections)*	0.628
Individual functional connections (× 20 Runs)	
100 Random	0.455
Whole-brain (MLP)	0.495
Whole-brain (GCN)	0.483
Top 5 Cerebellar-prefrontal*	0.735
Top 5 whole-brain*	0.742

Table 2: Details about individuals used in this study. We group participants with PD diagnosis by their MDS-UPDRS 3.10 score (gait impairment severity). Walking time represents the total time an individual walked in the exam, double support time denotes the amount of time where both feet were on the ground, % double percentage represents the percentage of the time walking where the participant had both feet on the ground, and average torso angle denotes the angle of the vector from the participant's spine to their neck from the ground. The Healthy/Control (CTRL) group is used to additionally validate our subtyping approach (visualized in Supplementary Figure 6). The last column shows the difference between the PD and CTRL groups with respect to the extracted measurements from the gait videos. Comparisons are done using a two-sided t-test ($p < 0.05$). The $p-$values for Walking Time, Double Support Time, and Average Torso Angle are 2.86×10^{-10}, 0.001, and 1.79×10^{-4}, respectively. A full comparison of cohort distributions for extracted motion measurements is displayed in Extended Data Figure 4.

Diagnosis	PD				CTRL	PD vs. CTRL
Gait impairment score	All	1	2	3		
N (total = 54)	31	14	13	4	23	-
Sex (F / M)	13 / 18	4 / 10	7 / 6	2/2	12 / 11	-
Age (mean ± std)	69.0 ± 7.0	67.6 ± 6.2	68.4 ± 8.0	75.7 ± 1.1	67.9 ± 12.8	-
Walking Time (sec)	29.5 ± 8.9	28.9 ± 6.5	29.1 ± 9.7	32.6 ± 14.2	13.9 ± 4.3	PD > CTRL
Double Support Time (sec)	6.0 ± 4.3	6.7 ± 4.3	6.0 ± 4.9	3.5 ± 1.2	2.8 ± 1.7	PD > CTRL
% Double Support	19.8 ± 11.6	22.3 ± 11.9	19.7 ± 12.2	11.49 ± 4.3	19.2 ± 8.3	PD = CTRL
Average Torso Angle	83.9 ± 5.2	84.1 ± 6.0	85.0 ± 4.0	79.7 ± 4.7	88.7 ± 2.6	PD < CTRL

Figure Legends/Captions

Figure 1: Our data-driven Parkinson's disease subtyping using motion (from videos) and brain functional connectivity (from rs-fMRI). (A) Generating low dimensional features using Canonical Correlation Analysis (CCA) from motion and functional connectivity representations. (B) PD subtyping using the generated low-dimensional features. The plot shows the clustering of our subtypes in two dimensions using principal component analysis (PCA) of our representation space. (C) Visualization of clinical variables that exhibited statistically significant (p-value < 0.05) levels of expression between the three subtypes as measured by the Chi-square test for categorical variables and the Kruskal-Wallis test for continuous variables. Colors are generated according to the corresponding variable mean values of subtypes. The blue color signifies that the mean value of a variable for that subtype was higher than the mean across the whole cohort. The red color signifies that the average variable value for that subtype was lower than the mean across the whole cohort. A darker color denotes a larger difference between the subtype mean and the overall mean.

Figure 2: Subtype approach comparison. (A) Histogram of clustering scores (σ) for random cohort groupings in our motion/fMRI representation space with clinical variable-based subtyping, TD/PIGD, and our subtyping σ scores also plotted. A lower σ score signifies better clustering. Note that our subtypes were generated by minimizing this clustering score. (B) Visualization of how well fig:td/pigd subtypes and clinical variable-based subtypes are clustered in our subtyping space. CCX signifies the motion component of our space while CCY signifies the fMRI component. For CCX, each oval represents a single individual since there are multiple data points per individual. (C) Visualization of our subtype clustering.

Figure 3: Subtype (digital) biomarker discovery. Visualization of joint impact on (A) model ability to predict gait impairment and (B) subtype generation measured by subtype grouping similarity. We display joints in purple that have an impact score in the top 25%. The radius of the joint is proportional to the joint importance. For joints on the left and right sides, we take the maximum joint impact on either side to avoid discriminating between the left and right sides. For gait impairment prediction, we also calculate various motion metrics for the joints having the greatest effects in the model. Metric values underlined in green are for individuals without gait impairment (score 0, $N = 9$) and values underlined in orange are for individuals with gait impairment (scores 1-4, $N = 45$). Metrics with a significant difference between individuals with and without gait impairment (two-sided p-value < 0.05 from t-test) are displayed. For (B), we also visualize the impact of fMRI connections, where the width of the fMRI connection is proportional to the connection's impact on subtype grouping. Nodes are colored according to subnetwork region.

References

[1] David Arthur and Sergei Vassilvitskii. "K-means++ the advantages of careful seeding". In: *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. 2007, pp. 1027–1035.

[2] Silvia Basaia et al. "Cerebro-cerebellar motor networks in clinical subtypes of Parkinson's disease". In: *npj Parkinson's Disease* 8.1 (2022), p. 113.

[3] Daniele Belvisi et al. "The pathophysiological correlates of Parkinson's disease clinical subtypes". In: *Movement Disorders* 36.2 (2021), pp. 370–379.

[4] Ralph HB Benedict et al. "Revision of the Brief Visuospatial Memory Test: Studies of normal performance, reliability, and validity." In: *Psychological assessment* 8.2 (1996), p. 145.

[5] Arthur L Benton, Nils R Varney, and Kerry deS Hamsher. "Visuospatial judgment: A clinical test". In: *Archives of neurology* 35.6 (1978), pp. 364–367.

[6] Christopher R Bowie and Philip D Harvey. "Administration and interpretation of the Trail Making Test". In: *Nature protocols* 1.5 (2006), pp. 2277–2281.

[7] Matthew Brendel et al. "Comprehensive subtyping of Parkinson's disease patients with similarity fusion: A case study with BioFIND data". In: *npj Parkinson's Disease* 7.1 (2021), p. 83.

[8] Christopher Alan Brown. *The Default Mode Network and Executive Function: Influence of Age, White Matter Connectivity, and Alzheimer's Pathology*. University of Kentucky, 2017.

[9] Mark Ming-Tso Chiang and Boris Mirkin. "Intelligent choice of the number of clusters in k-means clustering: an experimental study with different cluster spreads". In: *Journal of classification* 27 (2010), pp. 3–40.

[10] Julien Crémers et al. "Brain activation pattern related to gait disturbances in Parkinson's disease". In: *Movement Disorders* 27.12 (2012), pp. 1498–1505.

[11] Rachel A Crockett et al. "Resting state default mode network connectivity, dual task performance, gait speed, and postural sway in older adults with mild cognitive impairment". In: *Frontiers in aging neuroscience* 9 (2017), p. 423.

[12] Dean C Delis et al. "California Verbal Learning Test–Second Edition". In: *Assessment* (2000).

[13] EA Disbrow et al. "Resting state functional connectivity is associated with cognitive dysfunction in non-demented people with Parkinson's disease". In: *Journal of Parkinson's disease* 4.3 (2014), pp. 453–465.

[14] Robert S. Eisinger et al. "Motor subtype changes in early Parkinson's disease". In: *Parkinsonism & Related Disorders* 43 (2017), pp. 67–72. ISSN: 1353-8020. DOI: https://doi.org/10.1016/j.parkreldis.2017.07.018. URL: https://www.sciencedirect.com/science/article/pii/S1353802017302651.

[15] Mark Endo et al. "GaitForeMer: Self-Supervised Pre-Training of Transformers via Human Motion Forecasting for Few-Shot Gait Impairment Severity Estimation". In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VIII*. Springer. 2022, pp. 130–139.

[16] Mark Endo et al. *markendo/PD-Subtyping-Motion-Brain*. Zenodo. 2024. DOI: https://doi.org/10.5281/zenodo.12555317.

[17] Thomas Foltynie, Carol Brayne, and Roger A Barker. "The heterogeneity of idiopathic Parkinson's disease". In: *Journal of neurology* 249 (2002), pp. 138–145.

[18] I. Frommann et al. "Distinct patterns of cognitive impairment in multiple system atrophy patients of cerebellar and parkinsonian predominance". In: *Basal Ganglia* 2.2 (2012), pp. 91–96. ISSN: 2210-5336. DOI: https://doi.org/10.1016/j.baga.2012.02.001. URL: https://www.sciencedirect.com/science/article/pii/S2210533612000172.

[19] Christopher G Goetz et al. "Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results". In: *Movement disorders: official journal of the Movement Disorder Society* 23.15 (2008), pp. 2129–2170.

[20] Margot Heijmans et al. "Monitoring Parkinson's disease symptoms during daily life: a feasibility study". In: *npj Parkinson's Disease* 5.1 (2019), p. 21.

[21] Jarmo Heinonen et al. "Default mode and executive networks areas: association with the serial order in divergent thinking". In: *PloS one* 11.9 (2016), e0162234.

[22] BS Hoffland et al. “A gait paradigm reveals different patterns of abnormal cerebellar motor learning in primary focal dystonias”. In: *The Cerebellum* 13 (2014), pp. 760–766.

[23] Elston H Hooper. *Hooper visual organization test (VOT)*. 625 Alaska Avenue, Torrance, CA 90503-5124: Western Psychological Services, 1983.

[24] Yanbing Hou et al. “Default-mode network connectivity in cognitively unimpaired drug-naive patients with rigidity-dominant Parkinson’s disease”. In: *Journal of neurology* 264 (2017), pp. 152–160.

[25] Thorsten Joachims. “Transductive learning via spectral graph partitioning”. In: *Proceedings of the 20th international conference on machine learning (ICML-03)*. 2003, pp. 290–297.

[26] Edith Kaplan, Harold Goodglass, and Sandra Weintraub. “Boston naming test”. In: *The Clinical Neuropsychologist* (1983).

[27] Prasanna R. Karunanayaka et al. “Default mode network differences between rigidity- and tremor-predominant Parkinson’s disease”. In: *Cortex* 81 (2016), pp. 239–250. ISSN: 0010-9452. DOI: https://doi.org/10.1016/j.cortex.2016.04.021. URL: https://www.sciencedirect.com/science/article/pii/S0010945216301046.

[28] Laurie A King et al. “Cognitively challenging agility boot camp program for freezing of gait in Parkinson disease”. In: *Neurorehabilitation and neural repair* 34.5 (2020), pp. 417–427.

[29] Thomas N Kipf and Max Welling. “Semi-Supervised Classification with Graph Convolutional Networks”. In: *International Conference on Learning Representations*. 2016.

[30] Muhammed Kocabas, Nikos Athanasiou, and Michael J Black. “Vibe: Video inference for human body pose and shape estimation”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020, pp. 5253–5263.

[31] Stuart Lloyd. “Least squares quantization in PCM”. In: *IEEE transactions on information theory* 28.2 (1982), pp. 129–137.

[32] Mandy Lu et al. “Quantifying Parkinson’s disease motor severity under uncertainty using MDS-UPDRS videos”. In: *Medical image analysis* 73 (2021), p. 102179.

[33] Eric McKimm. "Cerebellar Neuropathology Influences Cerebellar-Prefrontal Cortex Pathways: A Comprehensive Approach as Relevant to Autism Spectrum Disorders". PhD thesis. The University of Memphis, 2016.

[34] Raja Mehanna et al. "Comparing clinical features of young onset, middle onset and late onset Parkinson's disease". In: *Parkinsonism & Related Disorders* 20.5 (2014), pp. 530–534. ISSN: 1353-8020. DOI: https://doi.org/10.1016/j.parkreldis.2014.02.013. URL: https://www.sciencedirect.com/science/article/pii/S1353802014000662.

[35] Eva M Müller-Oehring et al. "Disruption of cerebellar-cortical functional connectivity predicts balance instability in alcohol use disorder". In: *Drug and alcohol dependence* 235 (2022), p. 109435.

[36] M Patricia Murray et al. "Walking patterns of men with parkinsonism". In: *American Journal of Physical Medicine & Rehabilitation* 57.6 (1978), pp. 278–294.

[37] Ziad S Nasreddine et al. "The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment". In: *Journal of the American Geriatrics Society* 53.4 (2005), pp. 695–699.

[38] Favour Nerrise et al. "An Explainable Geometric-Weighted Graph Attention Network for Identifying Functional Networks Associated with Gait Impairment". In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2023: 26th International Conference, Vancouver, Canada, October 8-12, 2023, Proceedings*. Springer. 2023.

[39] Claire O'Callaghan et al. "Cerebellar atrophy in Parkinson's disease and its implication for network connectivity". In: *Brain* 139.3 (2016), pp. 845–855.

[40] Gennaro Pagano et al. "Age at onset and Parkinson disease phenotype". In: *Neurology* 86.15 (2016), pp. 1400–1407.

[41] Bart Post et al. "Clinical heterogeneity in newly diagnosed Parkinson's disease". In: *Journal of neurology* 255 (2008), pp. 716–722.

[42] Manuel Rodriguez et al. "Parkinson's disease as a result of aging". In: *Aging cell* 14.3 (2015), pp. 293–308.

[43] Adrià Rofes et al. "What drives task performance during animal fluency in people with Alzheimer's disease?" In: *Frontiers in psychology* 11 (2020), p. 1485.

[44] Marianna Selikhova et al. “A clinico-pathological study of subtypes in Parkinson’s disease”. In: *Brain* 132.11 (2009), pp. 2947–2957.

[45] Amir Shahroudy et al. “Ntu rgb+ d: A large scale dataset for 3d human activity analysis”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 1010–1019.

[46] Aaron Smith. “Symbol Digit Modalities Test”. In: *The Clinical Neuropsychologist* (1973).

[47] Otfried Spreen. “Neurosensory center comprehensive examination for aphasia”. In: *Neuropsychological laboratory* (1977).

[48] J Ridley Stroop. “Studies of interference in serial verbal reactions.” In: *Journal of experimental psychology* 18.6 (1935), p. 643.

[49] Alessandro Tessitore et al. “Resting-state brain connectivity in patients with Parkinson’s disease and freezing of gait”. In: *Parkinsonism & related disorders* 18.6 (2012), pp. 781–787.

[50] Mary Ann Thenganatt and Joseph Jankovic. “Parkinson disease subtypes”. In: *JAMA neurology* 71.4 (2014), pp. 499–504.

[51] Bruce Thompson. *Canonical correlation analysis: Uses and interpretation*. Vol. 47. Sage, 1984.

[52] Guadalupe J Torres et al. “A similarity measure for clustering and its applications”. In: *Int J Electr Comput Syst Eng* 3.3 (2009), pp. 164–170.

[53] Stephanie M Van Rooden et al. “The identification of Parkinson’s disease subtypes using cluster analysis: a systematic review”. In: *Movement disorders* 25.8 (2010), pp. 969–978.

[54] R. von Coelln et al. “The inconsistency and instability of Parkinson’s disease motor subtypes”. In: *Parkinsonism & Related Disorders* 88 (2021), pp. 13–18. ISSN: 1353-8020. DOI: https://doi.org/10.1016/j.parkreldis.2021.05.016. URL: https://www.sciencedirect.com/science/article/pii/S1353802021001875.

[55] Mary L Wagner et al. “Complications of disease and therapy: a comparison of younger and older patients with Parkinson’s disease”. In: *Annals of Clinical & Laboratory Science* 26.5 (1996), pp. 389–395.

[56] Doris D Wang and Julia T Choi. “Brain network oscillations during gait in Parkinson’s disease”. In: *Frontiers in human neuroscience* 14 (2020), p. 568703.

[57] Susan Whitfield-Gabrieli and Alfonso Nieto-Castanon. "Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks". In: *Brain connectivity* 2.3 (2012), pp. 125–141.

[58] Mirdhu M Wickremaratchi et al. "The motor phenotype of Parkinson's disease in relation to age at onset". In: *Movement Disorders* 26.3 (2011), pp. 457–463.

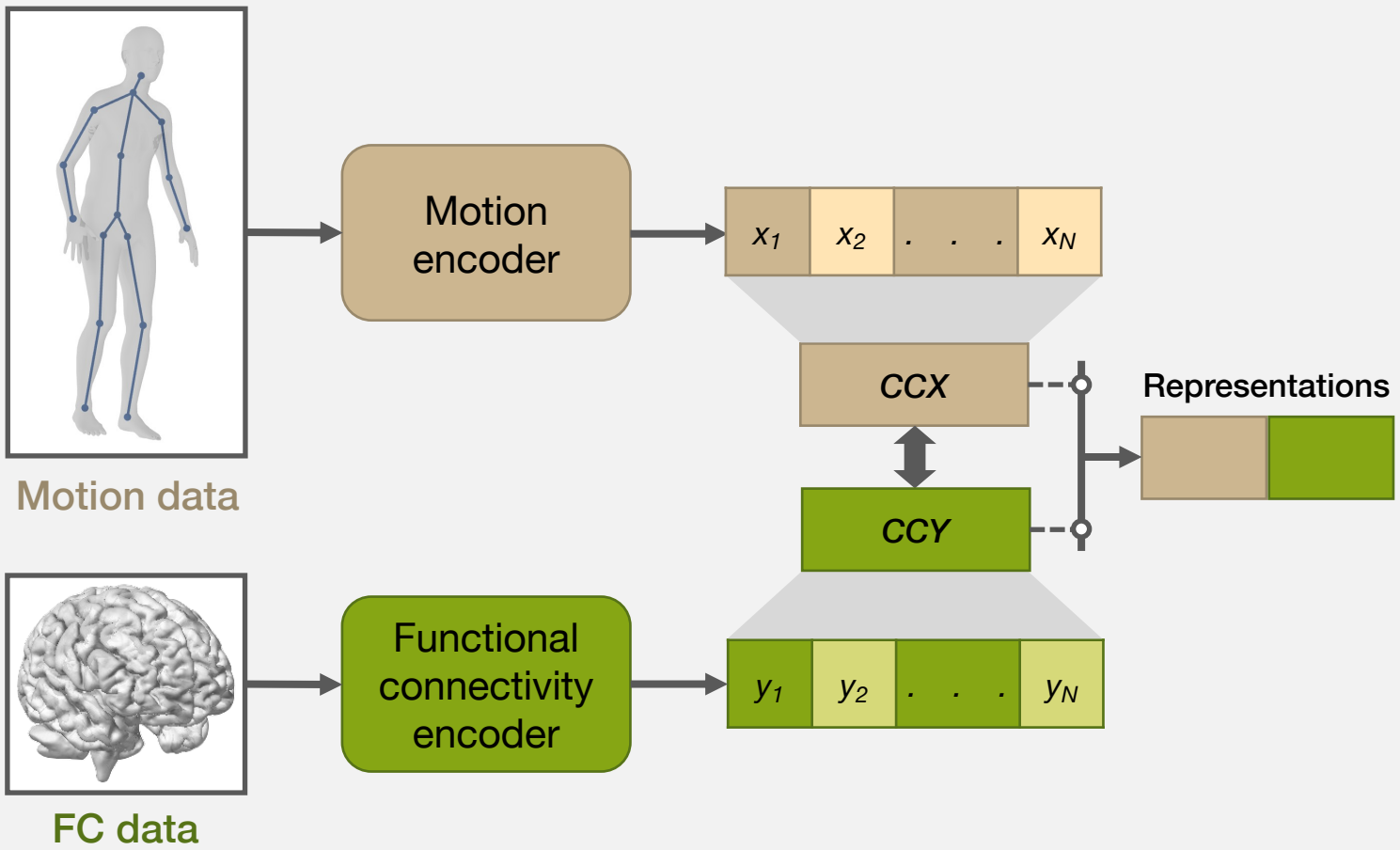
[59] Tao Wu and Mark Hallett. "The cerebellum in Parkinson's disease". In: *Brain* 136.3 (Feb. 2013), pp. 696–709.

[60] Xi Zhang et al. "Data-driven subtyping of Parkinson's disease using longitudinal clinical records: a cohort study". In: *Scientific reports* 9.1 (2019), p. 797.

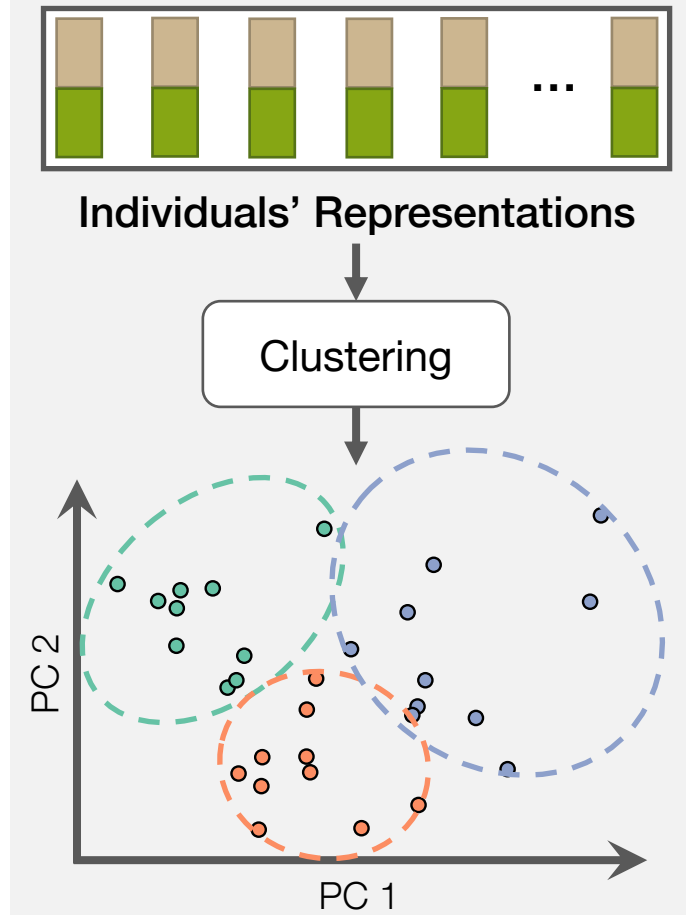
[61] Jianguo Zhong et al. "Levodopa imparts a normalizing effect on default-mode network connectivity in non-demented Parkinson's disease". In: *Neuroscience Letters* 705 (2019), pp. 159–166. ISSN: 0304-3940. DOI: https://doi.org/10.1016/j.neulet.2019.04.042. URL: https://www.sciencedirect.com/science/article/pii/S0304394019302848.

[62] Jill G. Zwicker et al. "Brain activation associated with motor skill practice in children with developmental coordination disorder: an fMRI study". In: *International Journal of Developmental Neuroscience* 29.2 (2011), pp. 145–152. ISSN: 0736-5748. DOI: https://doi.org/10.1016/j.ijdevneu.2010.12.002. URL: https://www.sciencedirect.com/science/article/pii/S0736574810004211.

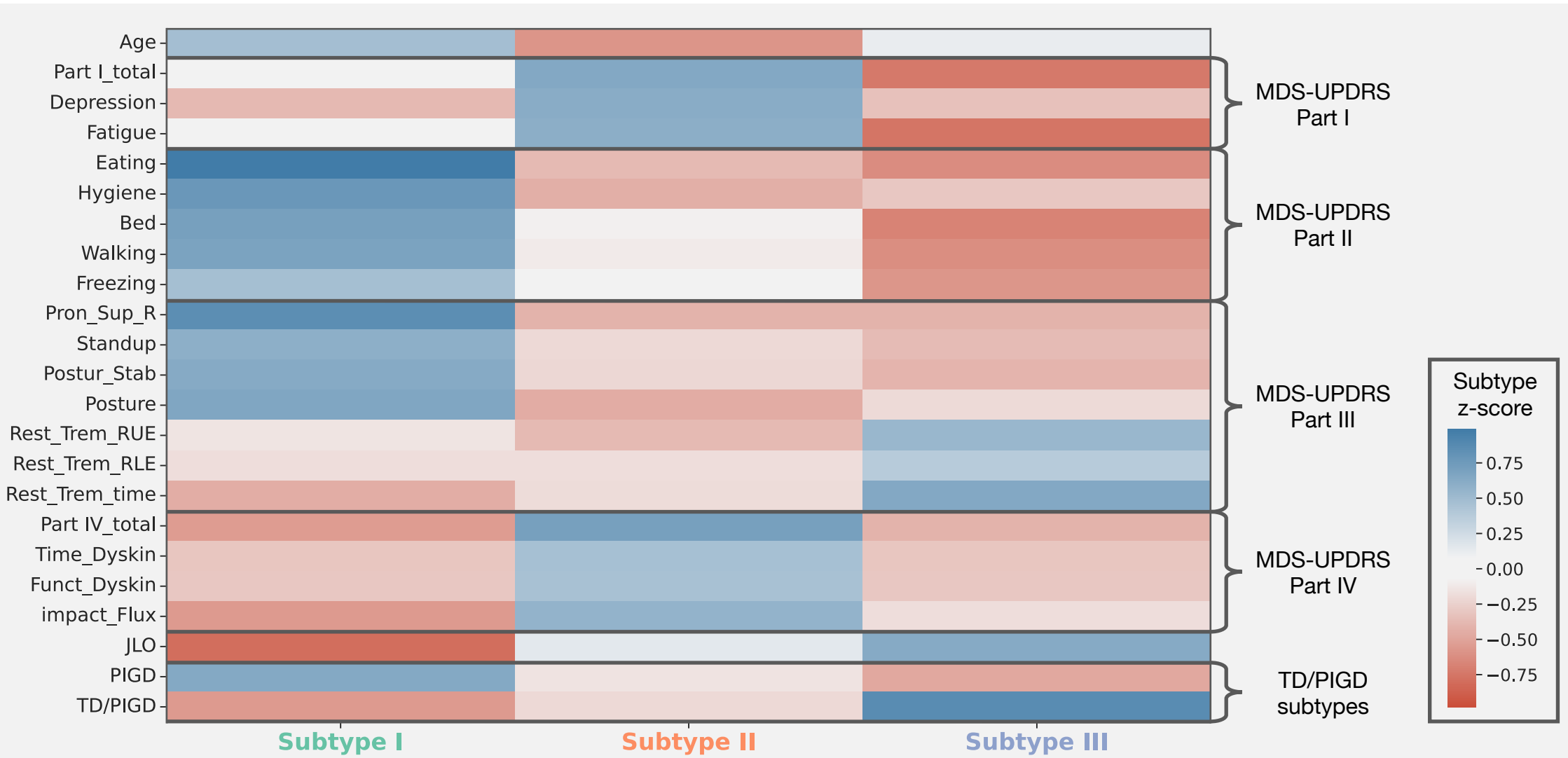
A Generating low-dimensional representations from motion and brain functional connectivity (FC) data



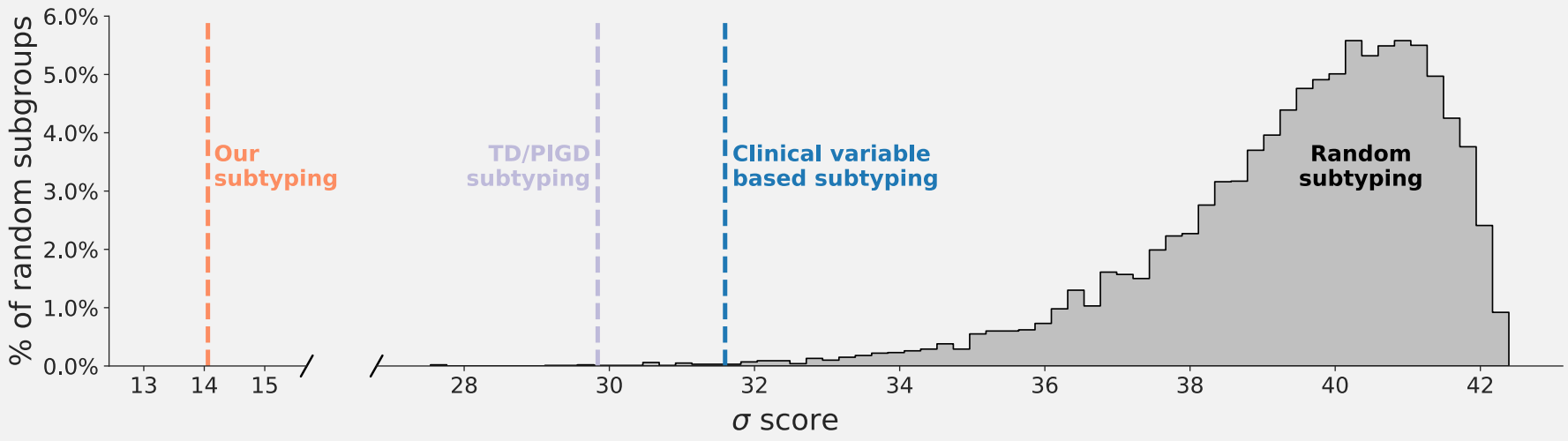
B Clustering based on motion-FC links



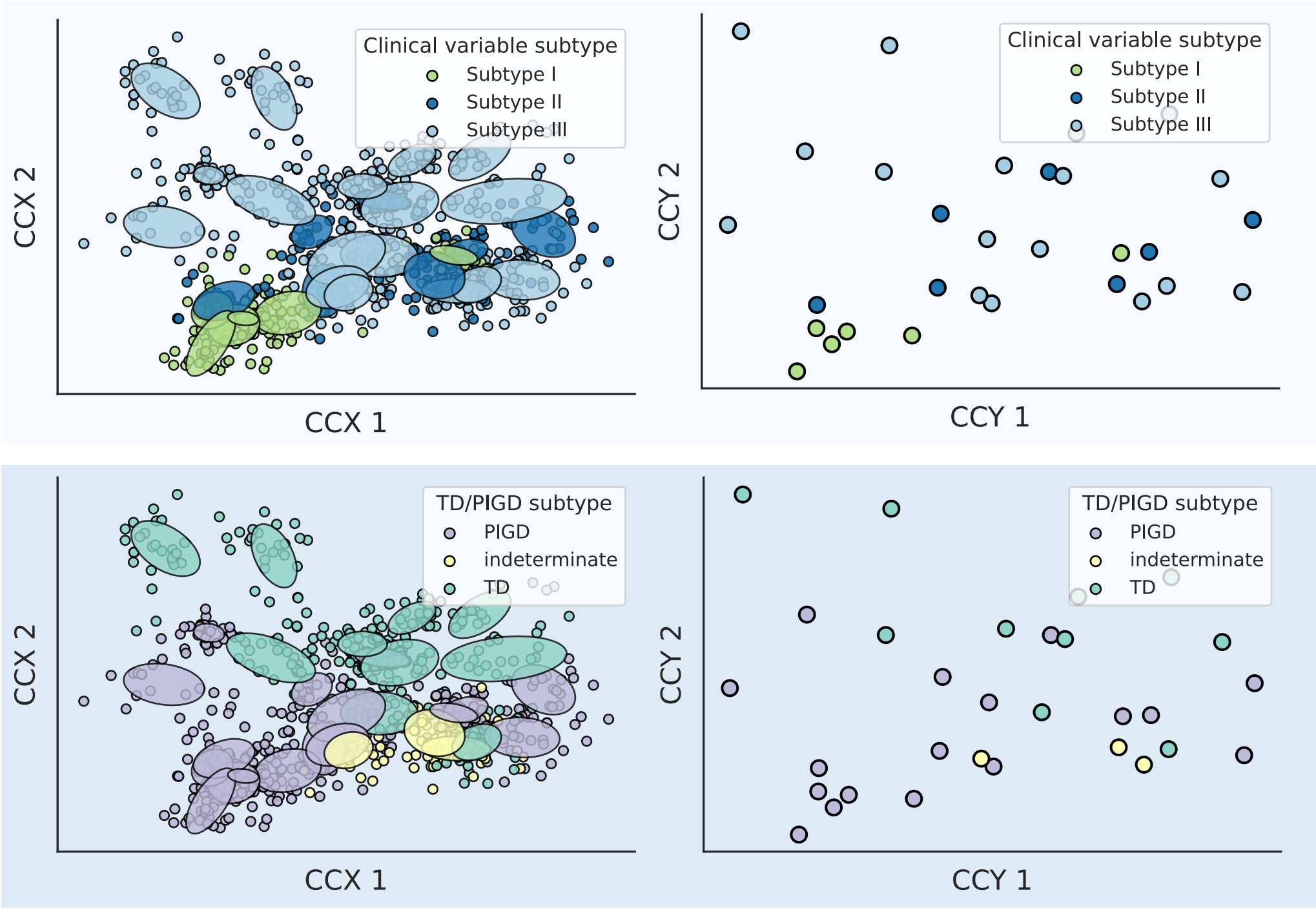
C Motion-FC subtype characteristics



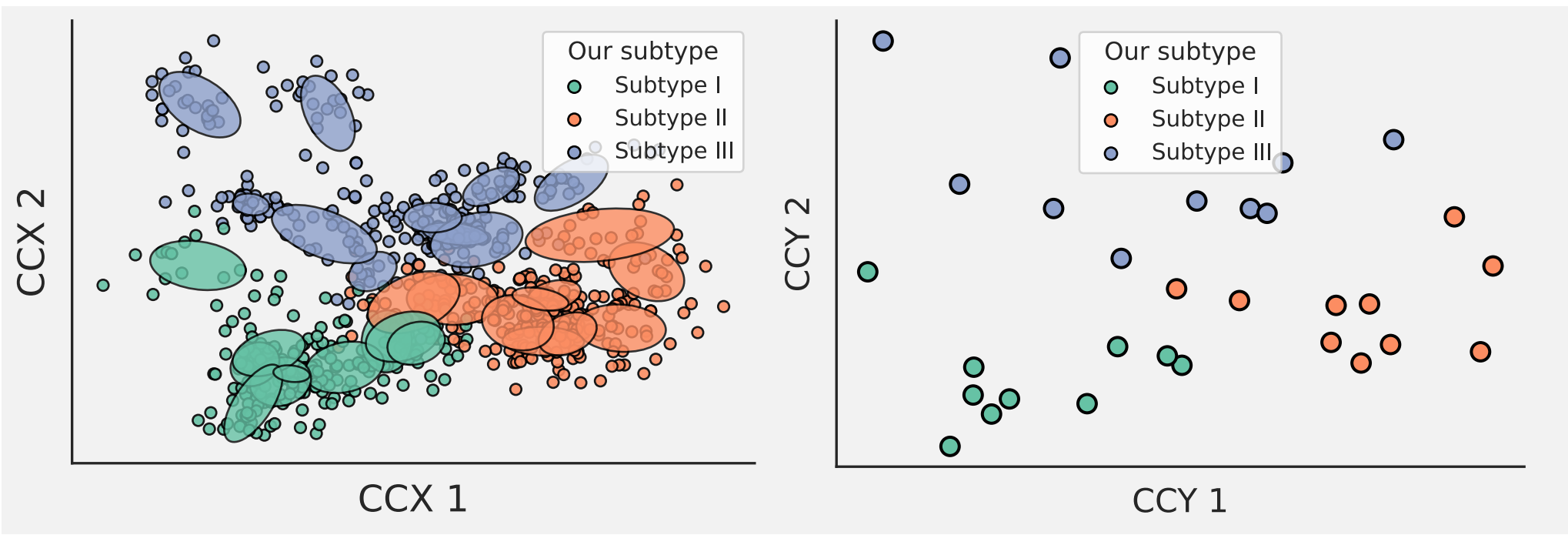
A Clustering score in our method's correlated representation space (σ) of various subtyping approaches

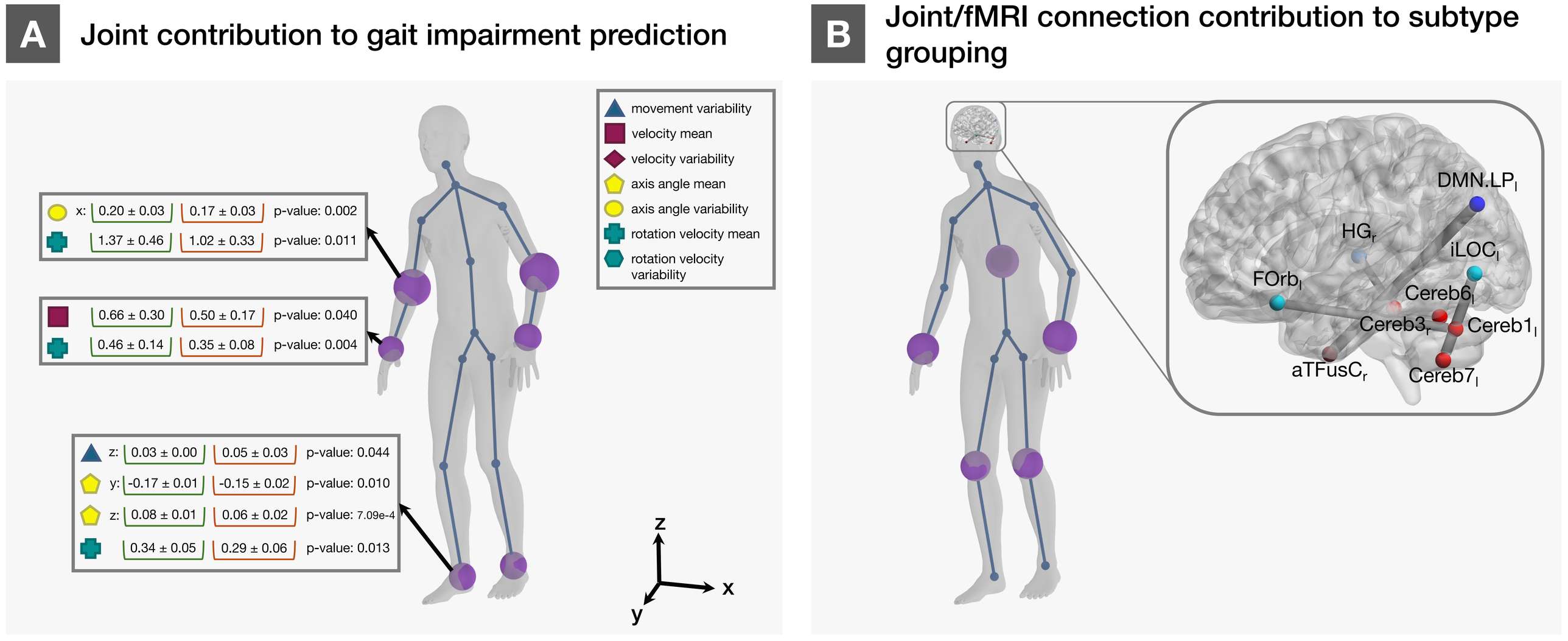


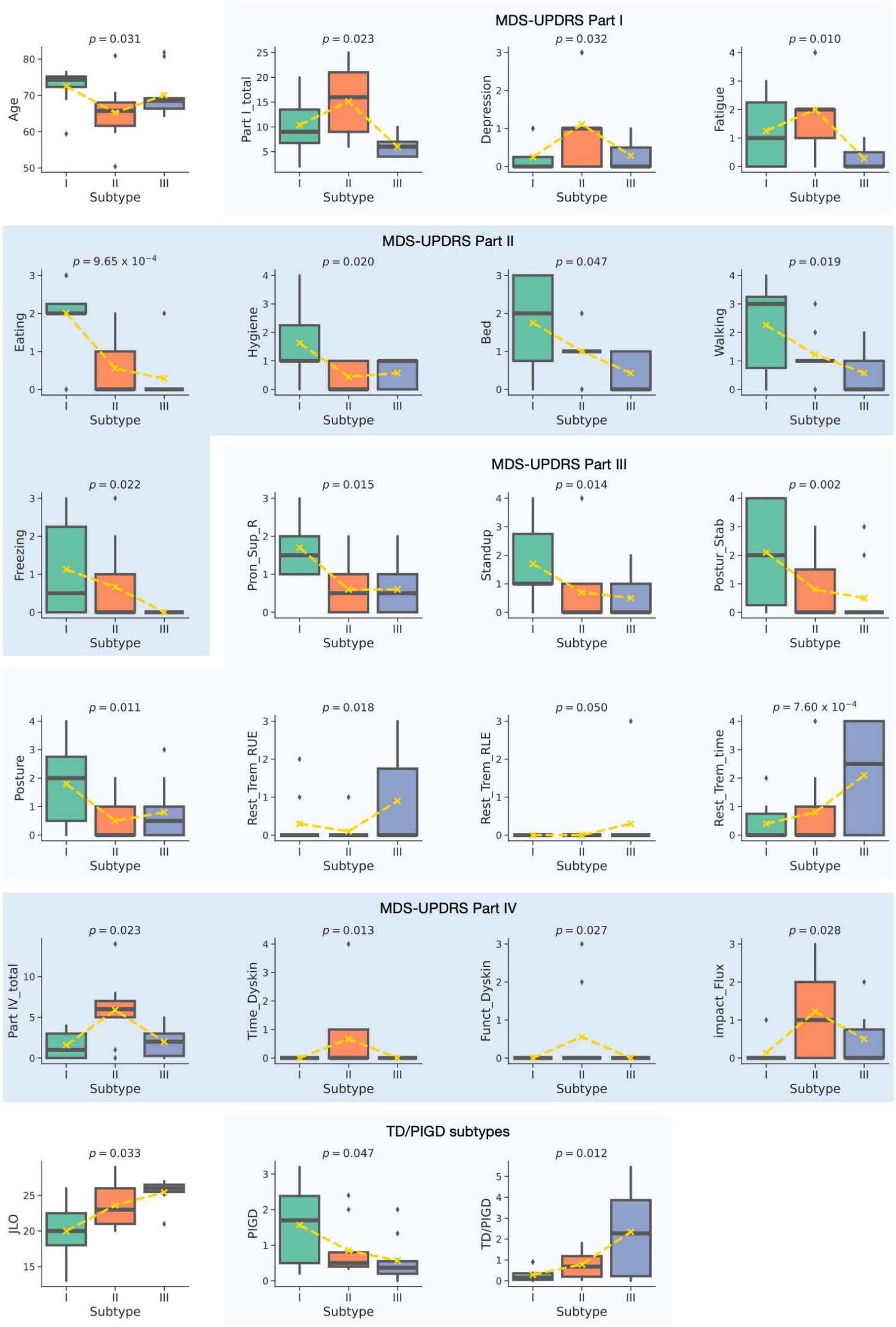
B Clustering of clinical variable-based subtypes and TD/PIGD subtypes in our subtyping space

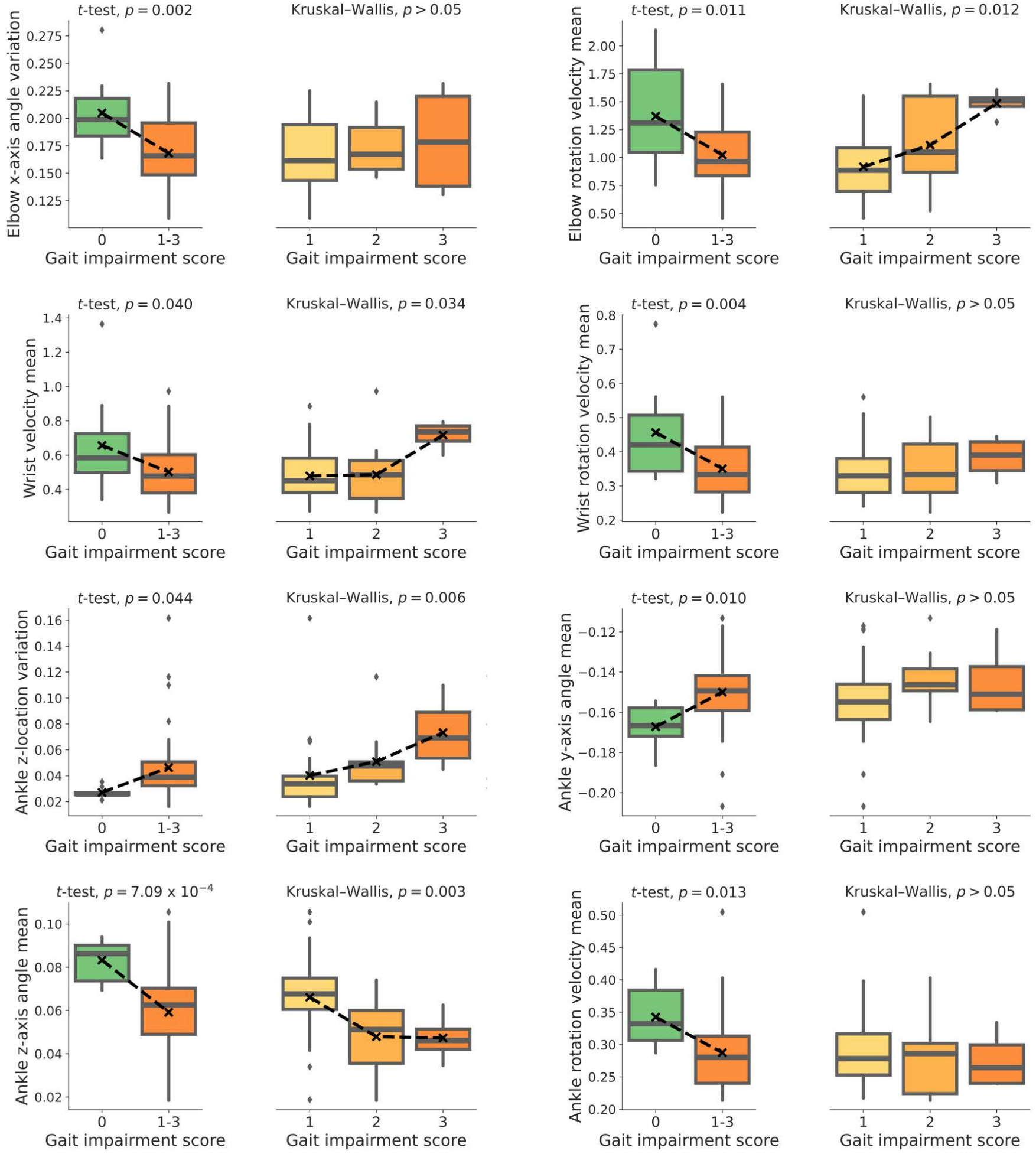


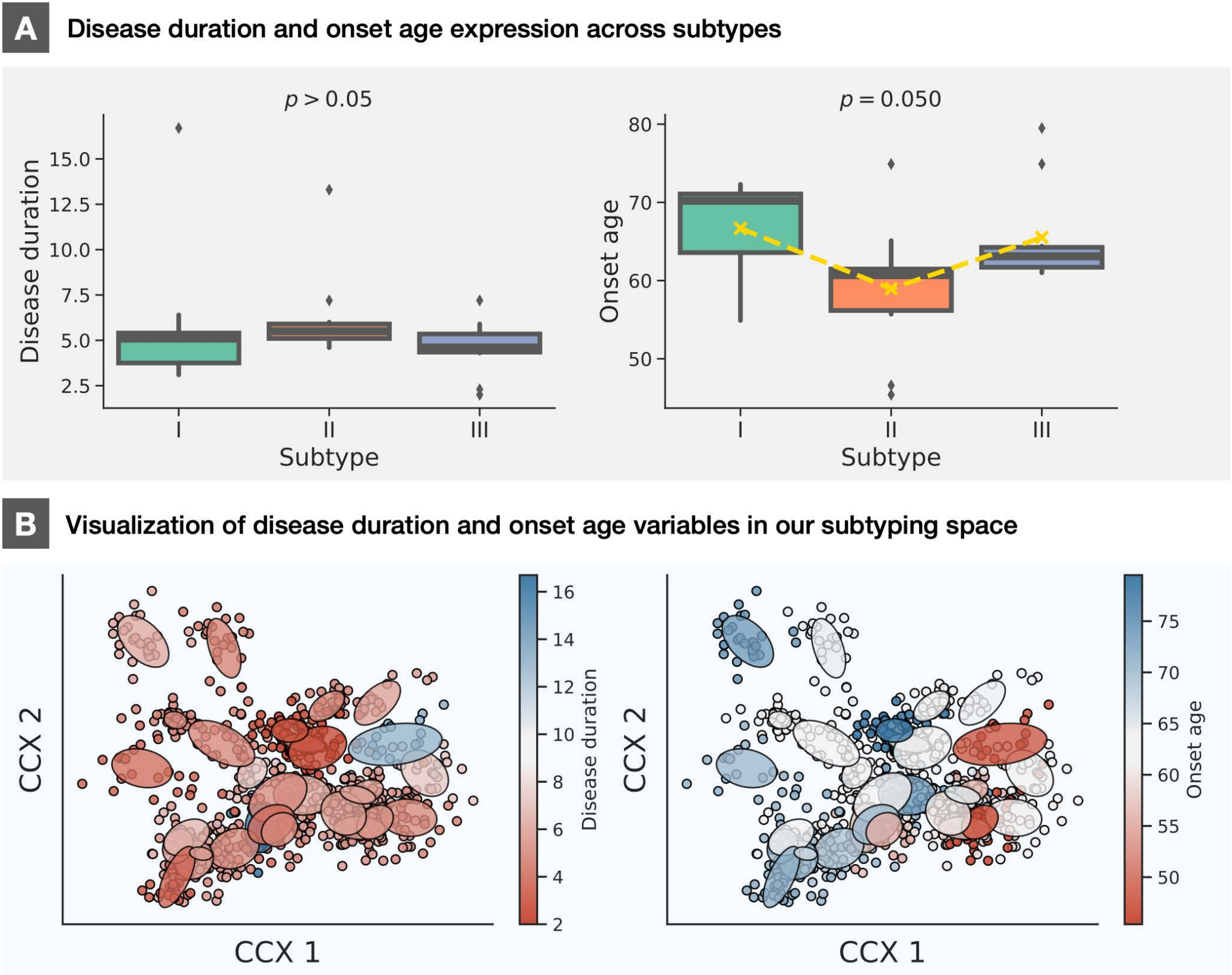
C Visualization of our subtype clustering

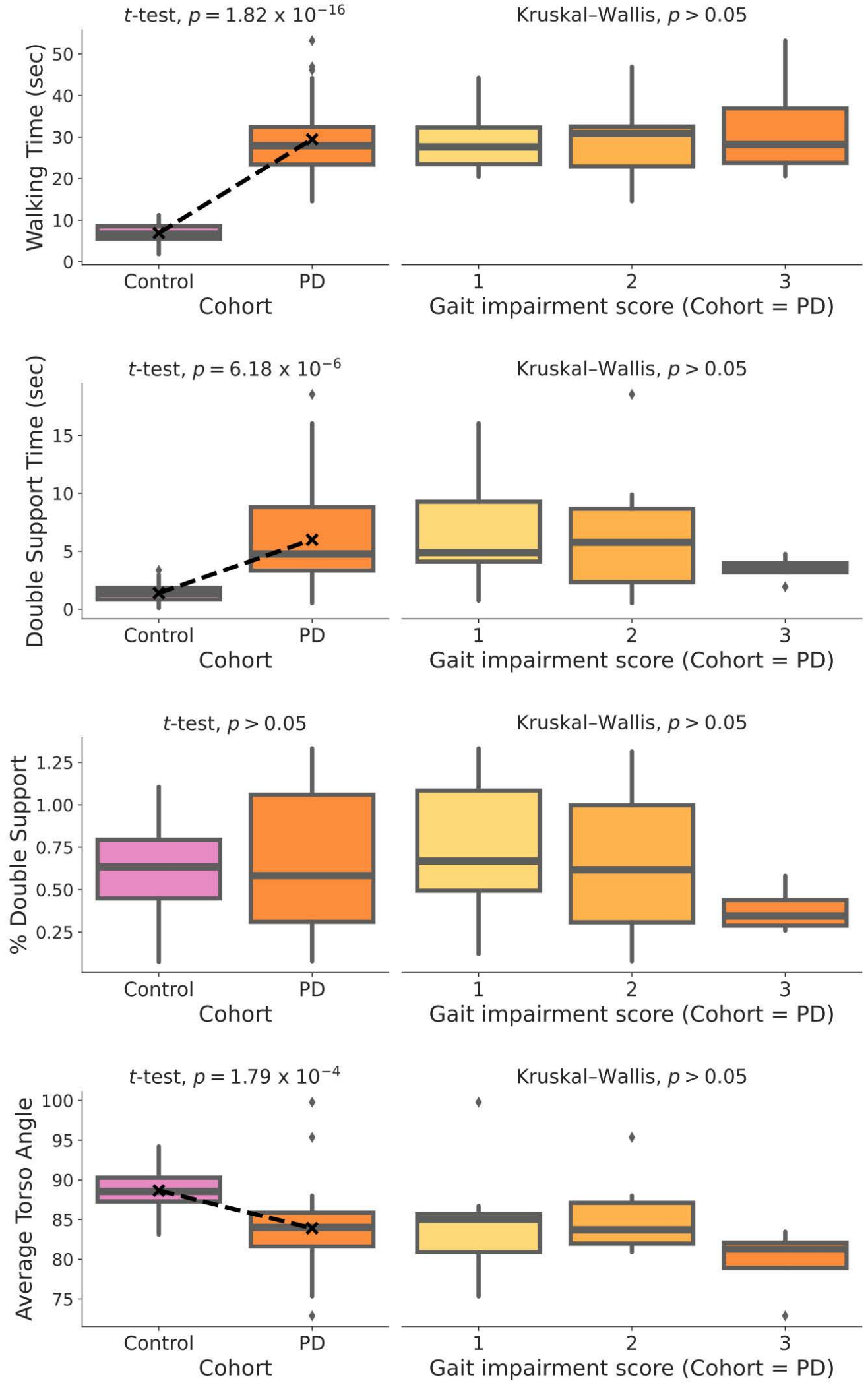



A
Joint contribution to gait impairment prediction
movement variability
velocity mean
velocity variability
axis angle mean
axis angle variability
rotation velocity mean
rotation velocity variability
x: 0.20 ± 0.03 0.17 ± 0.03 p-value: 0.002
1.37 ± 0.46 1.02 ± 0.33 p-value: 0.011
0.66 ± 0.30 0.50 ± 0.17 p-value: 0.040
0.46 ± 0.14 0.35 ± 0.08 p-value: 0.004
z: 0.03 ± 0.00 0.05 ± 0.03 p-value: 0.044
y: -0.17 ± 0.01 -0.15 ± 0.02 p-value: 0.010
z: 0.08 ± 0.01 0.06 ± 0.02 p-value: 7.09e-4
0.34 ± 0.05 0.29 ± 0.06 p-value: 0.013
z
x
y
B
Joint/fMRI connection contribution to subtype grouping
DMN.LP$_l$
HG$_r$
iLOC$_l$
FOrb$_l$
Cereb6$_l$
Cereb3$_r$
Cereb1$_l$
aTFusC$_r$
Cereb7$_l$


MDS-UPDRS Part I
p = 0.031
Age
p = 0.023
Part I_total
p = 0.032
Depression
p = 0.010
Fatigue
MDS-UPDRS Part II
p = 9.65 x 10−4
Eating
p = 0.020
Hygiene
p = 0.047
Bed
p = 0.019
Walking
MDS-UPDRS Part III
p = 0.022
Freezing
p = 0.015
Pron_Sup_R
p = 0.014
Standup
p = 0.002
Postur_Stab
p = 0.011
Posture
p = 0.018
Rest_Trem_RUE
p = 0.050
Rest_Trem_RLE
p = 7.60 x 10−4
Rest_Trem_time
MDS-UPDRS Part IV
p = 0.023
Part IV_total
p = 0.013
Time_Dyskin
p = 0.027
Funct_Dyskin
p = 0.028
impact_Flux
TD/PIGD subtypes
p = 0.033
JLO
p = 0.047
PIGD
p = 0.012
TD/PIGD
Subtype
I
II
III


t-test, p = 0.002
Kruskal–Wallis, p > 0.05
Elbow x-axis angle variation
t-test, p = 0.011
Kruskal–Wallis, p = 0.012
Elbow rotation velocity mean
t-test, p = 0.040
Kruskal–Wallis, p = 0.034
Wrist velocity mean
t-test, p = 0.004
Kruskal–Wallis, p > 0.05
Wrist rotation velocity mean
t-test, p = 0.044
Kruskal–Wallis, p = 0.006
Ankle z-location variation
t-test, p = 0.010
Kruskal–Wallis, p > 0.05
Ankle y-axis angle mean
t-test, p = 7.09 x 10−4
Kruskal–Wallis, p = 0.003
Ankle z-axis angle mean
t-test, p = 0.013
Kruskal–Wallis, p > 0.05
Ankle rotation velocity mean
0
1-3
1
2
3
Gait impairment score

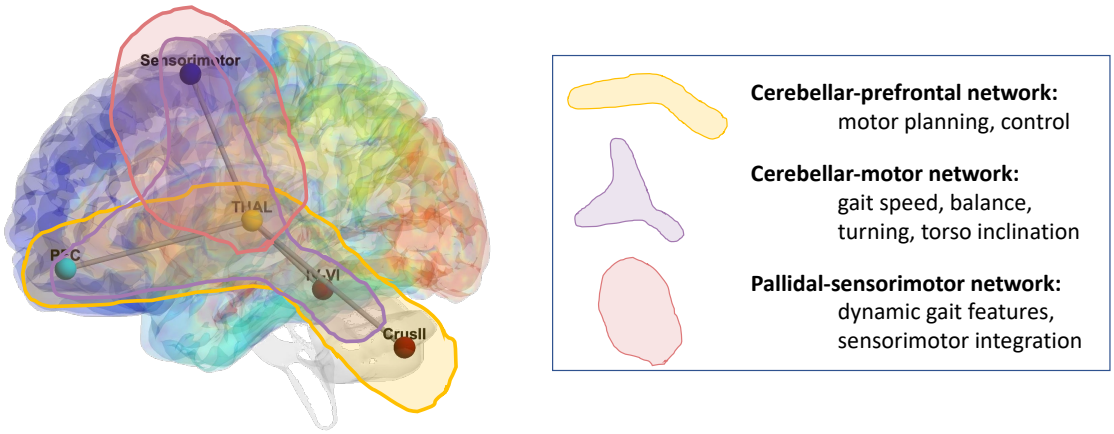

A
Disease duration and onset age expression across subtypes
p > 0.05
Disease duration
15.0
12.5
10.0
7.5
5.0
2.5
I
II
III
Subtype
p = 0.050
Onset age
80
70
60
50
I
II
III
Subtype
B
Visualization of disease duration and onset age variables in our subtyping space
CCX 2
CCX 1
16
14
12
10
8
6
4
2
Disease duration
CCX 2
CCX 1
75
70
65
60
55
50
Onset age


t-test, p = 1.82 x 10^−16
Kruskal–Wallis, p > 0.05
Walking Time (sec)
0
10
20
30
40
50
Control
PD
Cohort
1
2
3
Gait impairment score (Cohort = PD)
t-test, p = 6.18 x 10^−6
Kruskal–Wallis, p > 0.05
Double Support Time (sec)
0
5
10
15
Control
PD
Cohort
1
2
3
Gait impairment score (Cohort = PD)
t-test, p > 0.05
Kruskal–Wallis, p > 0.05
% Double Support
0.25
0.50
0.75
1.00
1.25
Control
PD
Cohort
1
2
3
Gait impairment score (Cohort = PD)
t-test, p = 1.79 x 10^−4
Kruskal–Wallis, p > 0.05
Average Torso Angle
75
80
85
90
95
100
Control
PD
Cohort
1
2
3
Gait impairment score (Cohort = PD)

Supplementary Information

Contents

1 Brain pathways' link to motor impairment



Supplementary Figure 1: Brain pathways related to motor outcome effects, including the cerebellar-prefrontal, cerebellar-motor, and pallidal-sensorimotor networks.

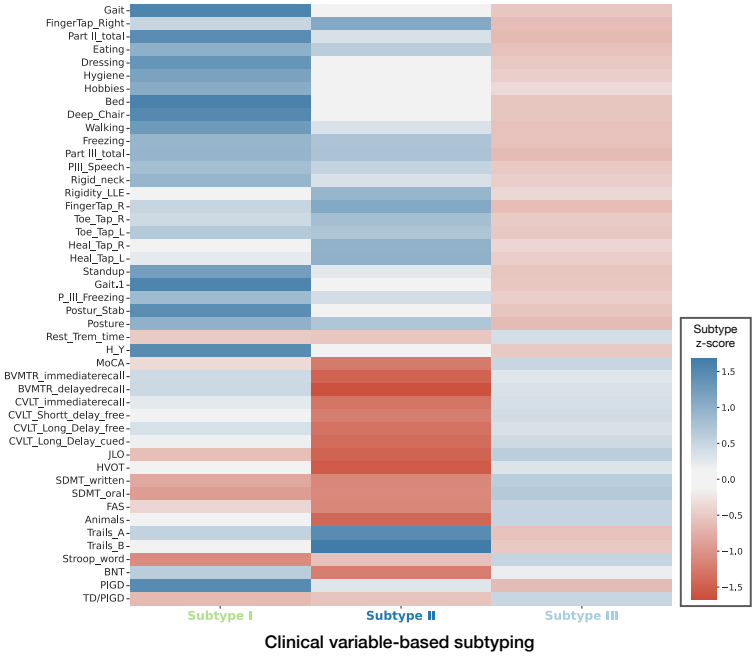
In Section 2.2, we identified and studied specific subnetworks that have been found to affect motor outcomes. The subnetworks we inspected were the cerebellar-prefrontal, cerebellar-motor, and pallidal-sensorimotor networks, identified based on the literature. Previous studies have uncovered cerebellar-prefrontal circuit involvement in various motor-related diseases and disorders. Ataxia in individuals with alcohol use disorder (AUD) is accompanied by cerebellar-frontal hyperconnectivity [6]. Cerebellar-prefrontal cortex pathways have shown an effect on autism spectrum disorders (ADSs) which are often accompanied by motor difficulties such as slow and repetitive hand and foot movements, slow and inaccurate dexterity, unstable balance, and impaired gait [5]. Cognitive impairments that correspond to the type of motor predominance in the cerebellar type of multiple system atrophy (MSA-C) reflect prevailing involvement of cerebellar-prefrontal circuits [3]. Children with developmental coordination disorder (DCD) demonstrate under-activation in the cerebellar–prefrontal network compared to typically developing peers [10]. Similarly, several previous studies have shown that the cerebellar-motor network has relevance to motor outcomes, especially for Parkinson's disease. One work found that compared to controls, PD individuals presented decreased stepwise functional connectivity in the sensorimotor network [1]. In another study, PD-related motor impairment severity was predicted by a combination of cerebellar atrophy and decreased cerebellar-sensorimotor connectivity [8]. Lastly, abnormalities in FC between pallidal-sensorimotor pathways in PD freezers compared to non-freezers have been reported [4].

2 Predicting motor impairment

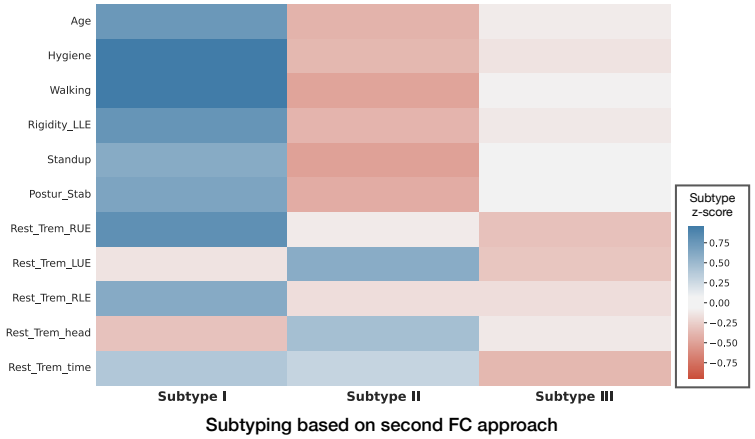
Supplementary Table 1: Predictive ability of gait impairment severity using either human motion (video) or FC (rs-fMRI) data. The model using human motion extracted skeleton and joint positions over time from videos. It used a motion encoder (based on our novel Transformer architecture [2]) to predict the gait impairment severity. The rs-fMRI top subnetwork approach used the five connections in the cerebellar-prefrontal network most correlated with gait impairment severity and used a simple multilayer perceptron (MLP) for prediction. The rs-fMRI top whole-brain approach input the five connections in the entire brain most correlated with gait impairment severity and used the same MLP for prediction. The rs-fMRI whole-brain approaches input all connections from the brain and either use an MLP or GCN for prediction. xGW-GAT is a weighted-graph attention neural network [7]. * indicates statistical difference by one-sided Wilcoxon signed-rank test at ($p < 0.05$) compared to a random classifier (area under the curve, AUC = 0.5) [9]. The $p-$values for the rs-fMRI - top subnetwork and rs-fMRI - top whole-brain are both 9.54×10^{-7}.

Approach	F_1 Score	Precision	Recall	AUC (one-vs-rest)
video - motion analysis [2]*	0.764	0.792	0.753	0.804
rs-fMRI - top subnetwork*	0.462	0.530	0.571	0.735
rs-fMRI - top whole-brain*	0.478	0.488	0.605	0.742
rs-fMRI - whole-brain - MLP	0.315	0.296	0.340	0.495
rs-fMRI - whole-brain - GCN	0.302	0.295	0.320	0.483
rs-fMRI - xGW-GAT [7]*	0.760	0.750	0.770	0.830

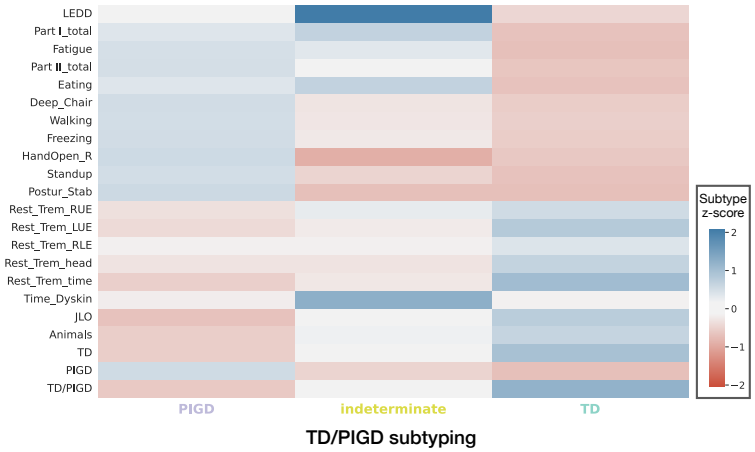
3 Comparison With Other Subtyping Frameworks



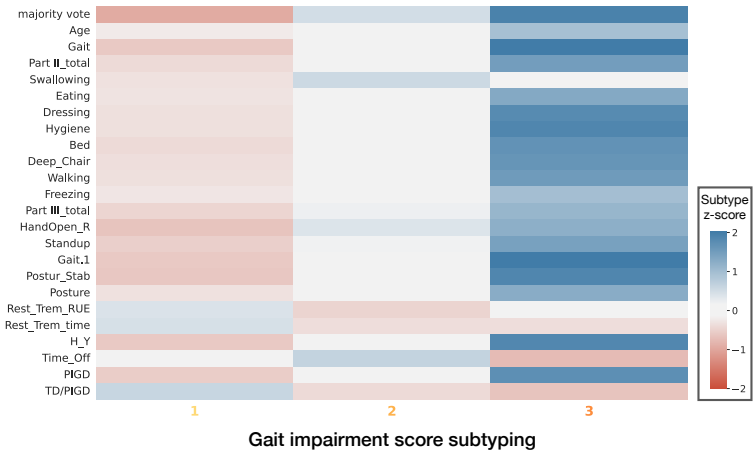
Supplementary Figure 2: Heatmap visualization of differentiated clinical variables from the conventional clinical variable-based subtyping.



Supplementary Figure 3: Heatmap visualization of differentiated clinical variables from subtyping based on the second FC selection approach explained in Section 2.2.

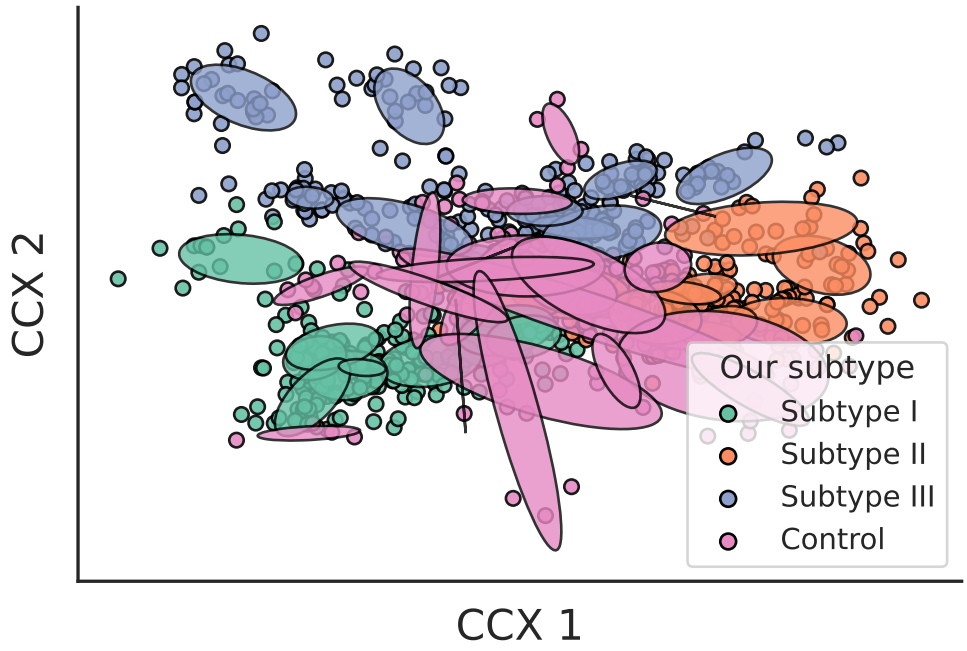


Supplementary Figure 4: Heatmap visualization of differentiated clinical variables from the TD/PIGD subtyping approach.



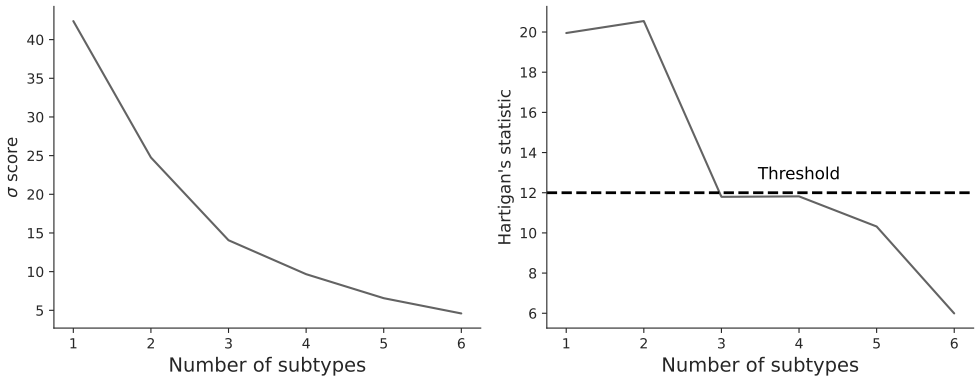
Supplementary Figure 5: Heatmap visualization of differentiated clinical variables from the subtyping based on gait impairment severity score.

4 Control Group Analysis



Supplementary Figure 6: Visualization of where controls (individuals without PD) are located in our subtype embedding space compared to other subtypes. We found that controls are clustered roughly together in the representation space between Subtype II and Subtype III. Its close proximity to Subtype II is logical as this subtype exhibited less overall symptom severity.

5 Number of Subtypes Selection



Supplementary Figure 7: Visualization of σ scores and Hartigan's statistic values as the number of subtypes increases. We found that the optimal number of subtypes is three when using a threshold of 12 for Hartigan's rule.

References

[1] Silvia Basaia et al. "Cerebro-cerebellar motor networks in clinical subtypes of Parkinson's disease". In: *npj Parkinson's Disease* 8.1 (2022), p. 113.

[2] Mark Endo et al. "GaitForeMer: Self-Supervised Pre-Training of Transformers via Human Motion Forecasting for Few-Shot Gait Impairment Severity Estimation". In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VIII*. Springer. 2022, pp. 130–139.

[3] I. Frommann et al. "Distinct patterns of cognitive impairment in multiple system atrophy patients of cerebellar and parkinsonian predominance". In: *Basal Ganglia* 2.2 (2012), pp. 91–96. ISSN: 2210-5336. DOI: https://doi.org/10.1016/j.baga.2012.02.001. URL: https://www.sciencedirect.com/science/article/pii/S2210533612000172.

[4] Laurie A King et al. "Cognitively challenging agility boot camp program for freezing of gait in Parkinson disease". In: *Neurorehabilitation and neural repair* 34.5 (2020), pp. 417–427.

[5] Eric McKimm. "Cerebellar Neuropathology Influences Cerebellar-Prefrontal Cortex Pathways: A Comprehensive Approach as Relevant to Autism Spectrum Disorders". PhD thesis. The University of Memphis, 2016.

[6] Eva M Müller-Oehring et al. "Disruption of cerebellar-cortical functional connectivity predicts balance instability in alcohol use disorder". In: *Drug and alcohol dependence* 235 (2022), p. 109435.

[7] Favour Nerrise et al. "An Explainable Geometric-Weighted Graph Attention Network for Identifying Functional Networks Associated with Gait Impairment". In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2023: 26th International Conference, Vancouver, Canada, October 8-12, 2023, Proceedings*. Springer. 2023.

[8] Claire O'Callaghan et al. "Cerebellar atrophy in Parkinson's disease and its implication for network connectivity". In: *Brain* 139.3 (2016), pp. 845–855.

[9] Frank Wilcoxon. *Individual comparisons by ranking methods*. One New York Plaza, Suite 4600, New York, NY 10004-1562: Springer, 1992.

[10] Jill G. Zwicker et al. "Brain activation associated with motor skill practice in children with developmental coordination disorder: an fMRI study". In: *International Journal of Developmental Neuroscience* 29.2 (2011), pp. 145–152. ISSN: 0736-5748. DOI: https://doi.org/10.1016/j.ijdevneu.2010.12.002. URL: https://www.sciencedirect.com/science/article/pii/S0736574810004211.