The accelerated aging model reveals critical mechanisms of late-onset Parkinson's disease

Shiyan Li[1], Hongxin Liu[1], Shiyu Bian[2], Xianzheng Sha[1], Yixue Li[3,4,5,6*] and Yin Wang[1,7*]

* Correspondence: yxli@sibs.ac.cn;
chinawangyin@foxmail.com
[3]Bio-Med Big Data Center, Key
Laboratory of Computational
Biology, CAS-MPG Partner Institute
for Computational Biology,
Shanghai Institute of Nutrition and
Health, Shanghai Institutes for
Biological Sciences, University of
Chinese Academy of Sciences,
Chinese Academy of Science,
Shanghai 200031, China
[1]Department of Biomedical
Engineering, School of
Fundamental Sciences, China
Medical University, Shenyang
110122, Liaoning Province, China
Full list of author information is
available at the end of the article

Abstract

Background: Late-onset Parkinson's disease (LOPD) is a common neurodegenerative disorder and lacks disease-modifying treatments, attracting major attentions as the aggravating trend of aging population. There were numerous evidences supported that accelerated aging was the primary risk factor for LOPD, thus pointed out that the mechanisms of PD should be revealed thoroughly based on aging acceleration. However, how PD was triggered by accelerated aging remained unclear and the systematic prediction model was needed to study the mechanisms of PD.

Results: In this paper, an improved PD predictor was presented by comparing with the normal aging process, and both aging and PD markers were identified herein using machine learning methods. Based on the aging scores, the aging acceleration network was constructed thereby, where the enrichment analysis shed light on key characteristics of LOPD. As a result, dysregulated energy metabolisms, the cell apoptosis, neuroinflammation and the ion imbalances were identified as crucial factors linking accelerated aging and PD coordinately, along with dysfunctions in the immune system.

Conclusions: In short, mechanisms between aging and LOPD were integrated by our computational pipeline.

Keywords: Late-onset Parkinson's disease, Aging acceleration, Network analysis

Introduction

As the second most common neurodegenerative syndrome (only after Alzheimer's disease), Parkinson's disease (PD, namely paralysis agitans), affects 1% of the worldwide population over the age of 60 years [1]. It is characterized by the motor (i.e. resting tremor, movement slowness, rigidity, and postural instability) and nonmotor symptoms (i.e. hyposmia, sleep disorders, autonomic dysfunction, neuropsychiatric alterations, and sensory symptoms) [1]. The motor and non-motor disorders caused by dopamine deficiency severely affect the life quality of the aging population. But up till now, there is not available curative and disease-

Li *et al. BioData Mining*   (2020) 13:4

Page 2 of 17

modifying therapies for PD [2]. With the prevalence of PD increasing, it is emerging as a socio-economic burden and a serious problem for the public health [2].

There were abundant epidemiological evidences supported that aging was the primary risk factor for Late-onset Parkinson's disease (LOPD) [3]. The prevalence of LOPD increased steadily with age [4], indicating aged brains were vulnerable to PD compared with young brains. Thus, the relationship between LOPD and aging was inseparable. First, there were some similar behavior changes between the healthy elderly people and the PD patients, such as slowness of motion, postural instability and cognitive deficits [5]. Second, the age at onset of PD significantly influenced the phenotype and progression of the disease [3]. In other words, age-related factors could predispose old individuals to develop this common neurodegenerative disease. Impaired mitochondrial functions [1, 5], the cell apoptosis [6, 7], and the chronic inflammatory process [8] showed commonalities in both PD and normal aged people. Third, the loss of substantia nigra (SN) neurons between normal aging and PD was significant, where the loss of neurons was 28.3% in older controls compared with younger controls, and the loss of pigmented neurons in PD was 73% compared with older controls [9]. It is noteworthy that SN neurons were the most susceptible to degeneration of PD [3]. Additionally, LOPD was often considered as a result of aging acceleration [5]. Therefore, to obtain a comprehensive understanding between aging and PD, we made a presumption that the onset of PD was induced by accelerated aging through a series of mechanisms (i.e. dysfunctions in the immune system, Fig. 1a).

Several studies have focused on the molecular mechanisms of PD development and in connection with the aging process [10]. Nevertheless, it is still insufficient to uncover the underlying relationship between PD and aging acceleration based on systematic models, which were informative to study PD. Fortunately, it has been demonstrated that machine learning techniques in correctly classifying healthy and PD patients performed well [11], thus predictive models based on genetic datasets were used to identify potential biomarkers, which was powerful to trial design and evaluation, as well as diagnosis and treatment of PD [12]. Recently, different approaches have been proposed for accurate characterization and diagnosis of PD (i.e. the voice disorder, gait, age and so on) by statistical analyses and machine learning methods [13, 14]. However, these results neither deeply revealed the connection between aging and PD, nor indicated the mechanisms of LOPD regulated by aging acceleration. Hence, systemic models were urgent to understand the trigger mechanisms of PD during aging acceleration (Fig. 1a), still needing to be improved by machine learning and network methods.

In the manuscript, to explore the disturbance of biological functions during the aging process, a computational pipeline was developed on the basis of both normal and PD transcriptional profiles (Fig. 1b). (1) The improved PD predictor (by comparing with normal aging) was contributed to identify risk biomarkers, as well as the aging predictor, where the kNN algorithm was used to capture data features from the complex non-linear characteristics of aging and neurodegenerative diseases, respectively [15]. (2) The aging scores were further summarized based on aging markers. (3) The aging accelerated network was constructed based
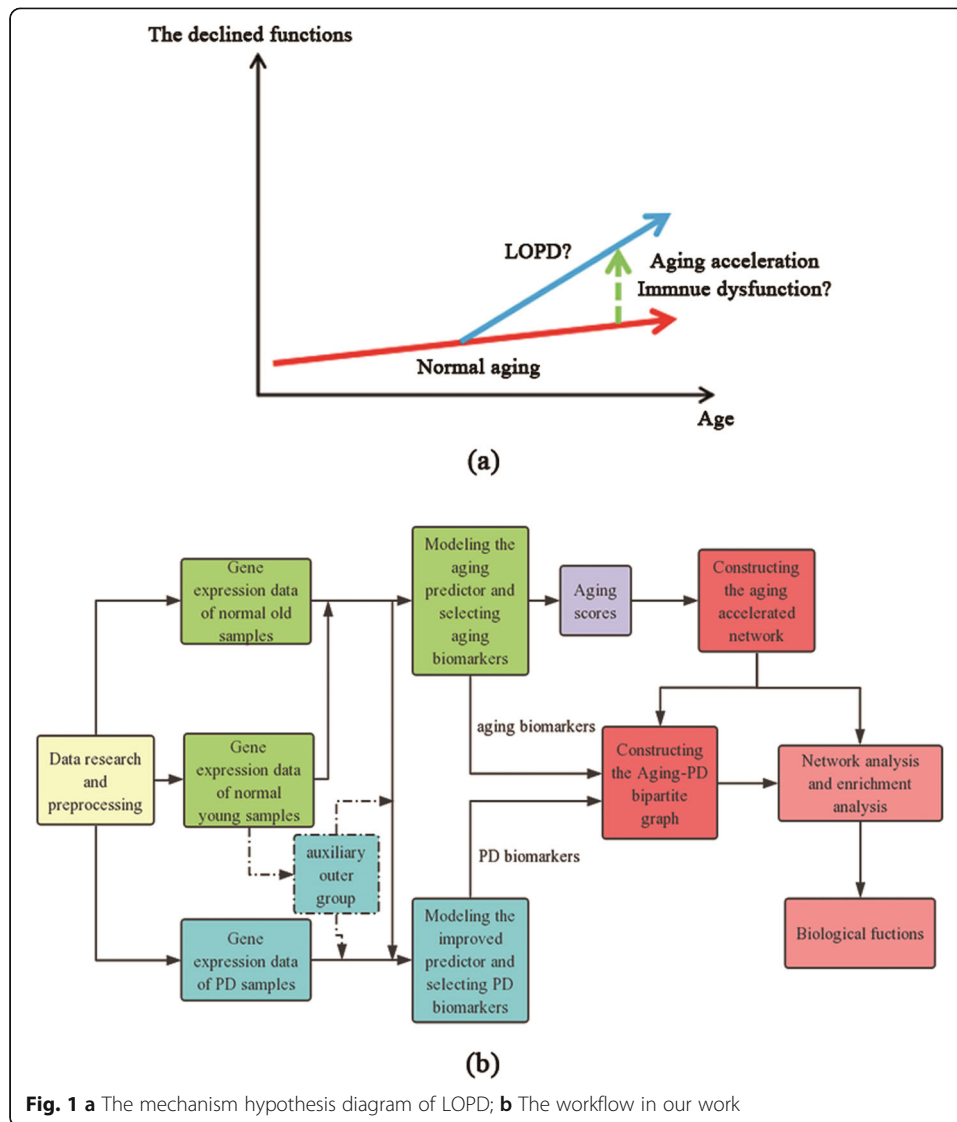
Fig. 1 a The mechanism hypothesis diagram of LOPD; **b** The workflow in our work

on both correlation and partial correlation of each gene pair. (4) Both network analysis and enrichment analysis were also conducted to identify potential mechanisms of LOPD regulated by aging acceleration. In conclusion, our pipeline was designed to investigate the connection between accelerated aging and LOPD at system level.

Results

Modeling the aging predictor and identifying aging markers

The total gene expression profile including 13,883 genes (Text S1) from six GSE profiles was analyzed. The details of the young samples and the old samples for the training dataset and test dataset were shown in Text S2. We utilized Pearson correlation coefficient and k-Nearest Neighbors (kNN, k = 5 and the cosine distance) to model the aging predictor and selected the optimal one based on the 10-fold cross-validation. The learning curve of the training dataset

was shown in Fig. 2a. As a result, the top 69 genes were identified as the aging markers (Table S1). The accuracy in the test dataset was 0.8109. The receiver operating characteristic (ROC) curve was shown in Fig. 2b, and the area under the ROC curve (AUC) was 0.81322. Obviously, these related results indicated the reliability of the optimal aging predictor.

It was noteworthy that the selected aging biomarkers were of great significance. For example, RBMS1 was the most relevant gene (with the highest ranking) among the 69 aging markers. As previously reported, RBMS1 encoded a cell cycle suppressor, which bound to an enhancer element of MYC [16]. The transcription protein of RBMS1 was implicated in DNA replication, transcription, and cell apoptosis [17]. As a result, these dysfunctional cell activities might lead to the accumulation of cellular senescence and aging [18]. In short, RBMS1 was closely related to the aging process.

Modeling the improved PD predictor and identifying PD markers

In order to better understand how LOPD was triggered in the context of aging acceleration, an improved PD predictor was constructed by discriminating PD and normal aged groups compared with the healthy young group. The details of the number of aged normal and PD samples in both training and test dataset were listed in Text S2.

Fig. 2 The learning curve and ROC curve in the aging predictor and improved PD predictor. **a** The learning curve of the aging predictor; **b** The ROC curve of the aging predictor; **c** The learning curve of the improved predictor; **d** The ROC curve of the improved PD predictor
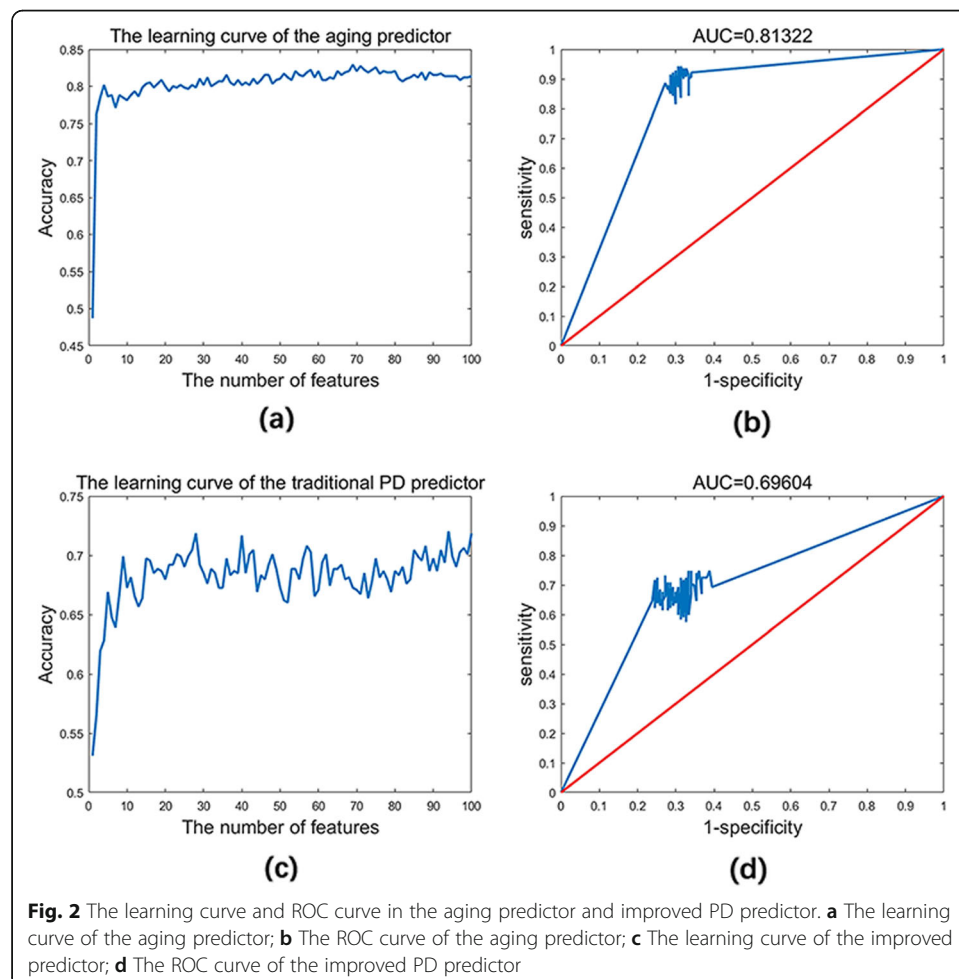
The importance of 13,883 genes was ranked by the absolute values of Pearson correlation coefficients, and the k-NN algorithm (k = 5 with the cosine distance) was used to model the PD predictor. Then the optimal model as well as PD markers were selected by the 10-fold cross-validation. The top 8 markers were considered as key PD markers (Table S2). And the learning curve was shown in Fig. 2c. For the stability of the selected model, it was validated in the total test dataset, the accuracy was 0.7039. The ROC curve for the PD predictor was shown in Fig. 2d and the AUC value was close to 0.7. To evaluate the accuracy of the improved PD predictor, we performed the same operation with the original gene profiles. The results showed that the prediction accuracy of the total test dataset was 0.6882 (Table 1). The related curves and biomarkers were also shown in Figure S1 and Table S3. In summary, compared with traditional PD predictor, the improved PD predictor performed more effectively.
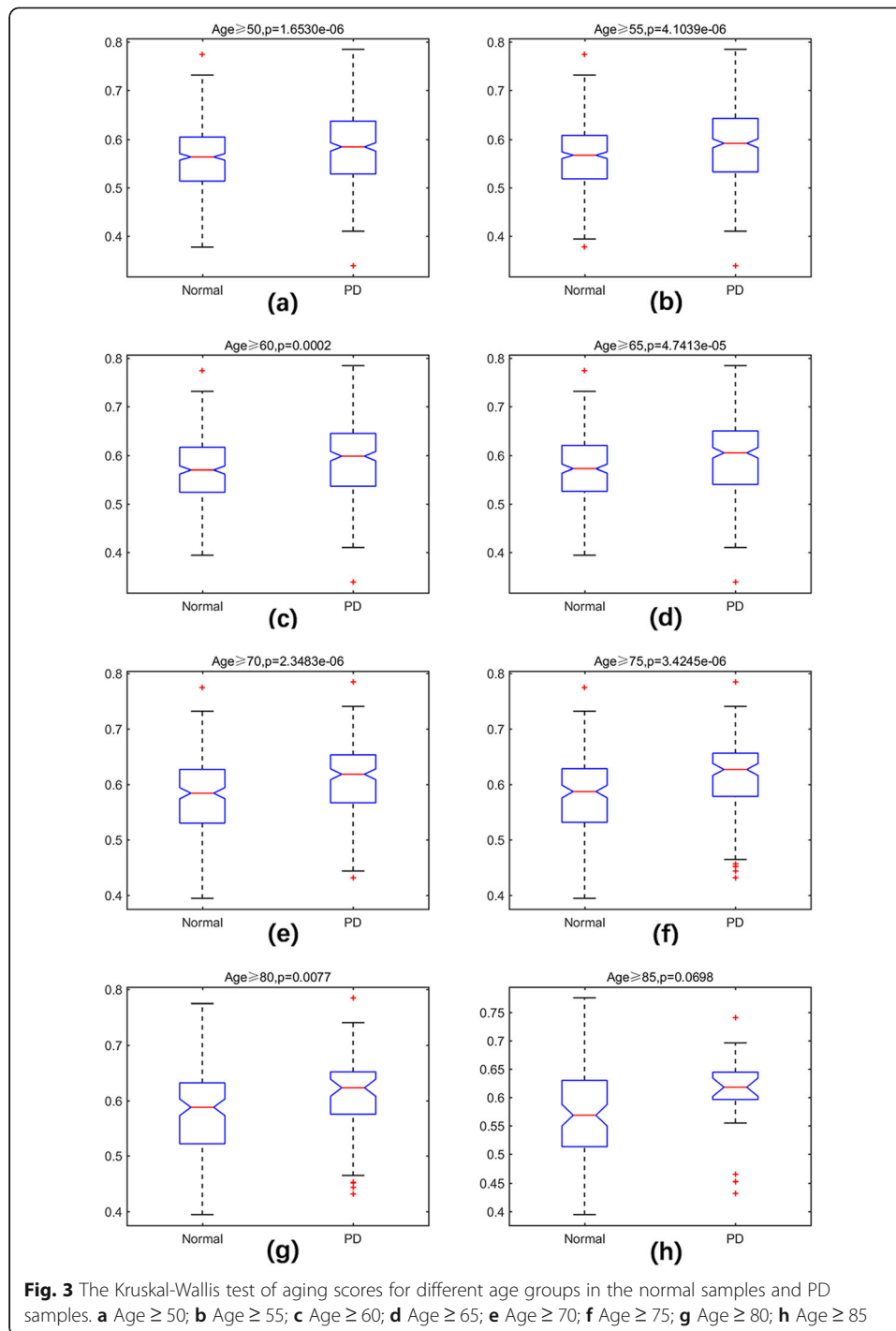
It was intriguing that ADD2 was identified as the top biomarker in the improved PD predictor. ADD2 was abundant in erythrocytes and brain tissue [19], and its proteins were associated with the assembly of spectrin-actin network and signal transduction pathways by interacting with protein kinase C-dependent and calcium/calmodulin-dependent pathways [20]. It has been reported that a delayed calcium efflux and mitochondrial Ca^{2+} overload may induce cell deaths and neuronal loss in neurodegenerative disorders [21]. Accordingly, ADD2 probably contributed to the PD pathology and was involved in the aging process by regulating the mitochondrial functions.

The aging acceleration pattern between PD and normal aged samples

To investigate the aging acceleration in LOPD, the aging scores were calculated by regressing the transformed age (using the sigmoid function) based on the aging markers. For distinct age groups, the median and mean of the chronological age and aging scores were shown in Table 2. There were increasing trends of the aging scores in both PD and normal aged samples, along with the chronological age. Apart from this, the aging scores of PD samples were always higher than normal samples within the same age group, although the median and mean of chronological age were lower. The results indicated that LOPD samples presented aging acceleration compared with normal aged samples. Therefore, when judging the level of aging, the chronological age did not fully reflect the aging rate of the body, but the aging score seemed to be more credible. In addition, the Kruskal-Wallis test was utilized to test the differences between LOPD and control samples. The results in Fig. 3 indicated that aging scores increased roughly with age, and almost all of the aging scores between PD and normal groups were significantly different ($p < 0.05$), revealing the accelerated aging patterns in most age groups of LOPD. Generally, the aging scores that used to evaluate the accelerated aging patterns in LOPD were of feasibility.

Table 1 The accuracy of the training dataset and test dataset for the aging predictor, the improved predictor as well as the traditional predictor

	The accuracy for training dataset	The accuracy for test dataset
The aging predictor	0.8291	0.8109
The improved PD predictor	0.7274	0.7039
The traditional PD predictor	0.7204	0.6882

Fig. 3 The Kruskal-Wallis test of aging scores for different age groups in the normal samples and PD samples. **a** Age ≥ 50; **b** Age ≥ 55; **c** Age ≥ 60; **d** Age ≥ 65; **e** Age ≥ 70; **f** Age ≥ 75; **g** Age ≥ 80; **h** Age ≥ 85

The aging acceleration network revealed potential biological functions

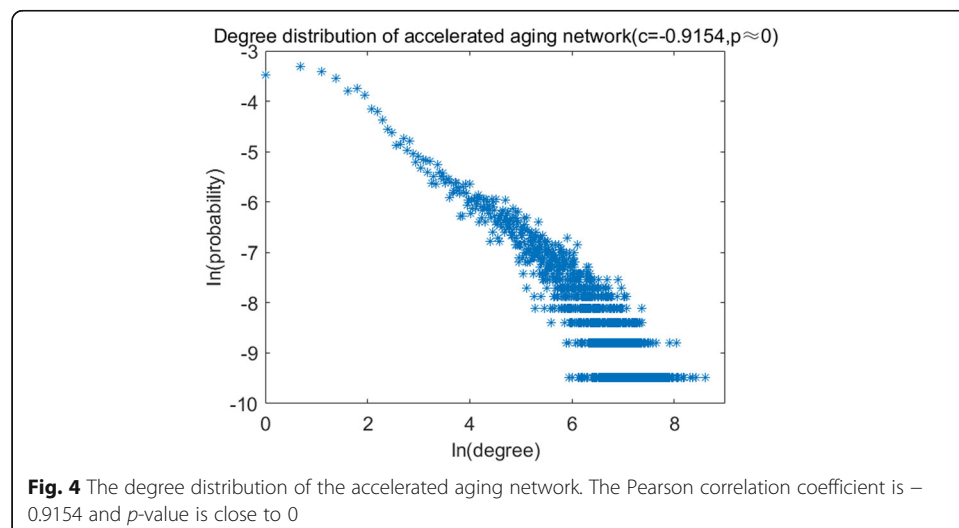To integrate aging and LOPD markers, the aging acceleration networks were constructed in the training data and the test data based on the aging scores (as mentioned in "Materials and Methods"), respectively. Further, the Fisher's exact test was carried out to calculate the similarity between these two networks, the result showed *p*-value was very closed to 0. The result indicated that our method was of reliability.

Table 2 The chronological age and aging scores of PD and control in different age groups

Age	The median age in PD	The median age in control	The mean age in PD	The mean age in control	The median aging score in PD	The median aging score in control	The mean aging score in PD	The mean aging score in control
≥50	69	71	68.82	71.21	0.584265	0.563570	0.582653	0.560505
≥55	70	72	70.39	72.65	0.591580	0.567050	0.585589	0.564299
≥60	72	76	72.43	77.02	0.598085	0.569840	0.588650	0.570399
≥65	75	79	74.79	79.84	0.604960	0.572590	0.594936	0.572388
≥70	76	82	77.20	82.30	0.617940	0.583710	0.606210	0.577525
≥75	78.5	85	79.90	85.26	0.626575	0.586700	0.614572	0.580568
≥80	84	87	83.68	88.02	0.622935	0.587725	0.610096	0.580391
≥85	87	90	86.77	90.57	0.617940	0.568660	0.603662	0.574285

To validate the scale-free characteristic, the curve of node degree distribution was shown in Fig. 4, where the logarithmic transformation was performed in both degrees and corresponding probabilities to test the power-law distribution. As a result, the Pearson correlation coefficient was − 0.9154 ($p \approx 0$). It illustrated that the aging acceleration network was in accordance with the scale-free characteristic (there was a negative correlation between the degree and the frequency, where only a small ratio of genes were with high degrees).

Considering key functions of the nodes with maximum node degrees, the gene nodes were ranked based on the degrees. SLC37A1 was with the maximum degree (=5465). SLC37A1 is a Pi-linked glucose-6-phosphate (G6P) antiporter and probably catalyzes both homologous (Pi/Pi) and heterologous (G6P/Pi) exchanges [22]. Furthermore, it has been reported that SLC37A1 is related to transport glycerol-3-phosphate (G3P), where the disturbance of ATP would affect these critical functions in brain [23]. In short, SLC37A1 may play an important role between aging and PD through the mitochondrial function.

Fig. 4 The degree distribution of the accelerated aging network. The Pearson correlation coefficient is − 0.9154 and *p*-value is close to 0

Li *et al. BioData Mining*   (2020) 13:4

Page 8 of 17

Underlying PD mechanisms based on the enrichment analysis in the aging acceleration network

To investigate the relationship between aging and LOPD, the Aging-PD shortest paths were identified based on aging acceleration network. Then the enrichment analysis (i.e. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and Biological Process (BP) terms, shown in Table 3) was performed based on each shortest path, respectively. The most significant KEGG pathway was the PPAR signaling pathway ($p = 2.0267e-05$, FDR = 0.0038, Fig. 5a). Interestingly, it was also with the maximum number of enriched paths (3 paths). There were three isotypes of PPAR: PPARα, PPARβ/δ and PPARγ, and their functions were related to the immunotolerance, lipid/glucose metabolism, angiogenesis and the inflammatory response [24]. PPARγ was critical to neuroprotective and anti-inflammatory responses, thus reduced inflammation-driven neuronal damages [25], cell deaths and the progression of neurodegeneration [26]. A current research even has indicated that the PPAR signaling pathway was as the potential target of neuroprotection [27]. Moreover, other results also showed that PPARγ was a critical therapeutic target in Parkinson's disease, by regulating fatty acid oxidation, immune responses and the mitochondrial function [28]. In short, the PPAR signaling pathway was vital to LOPD.
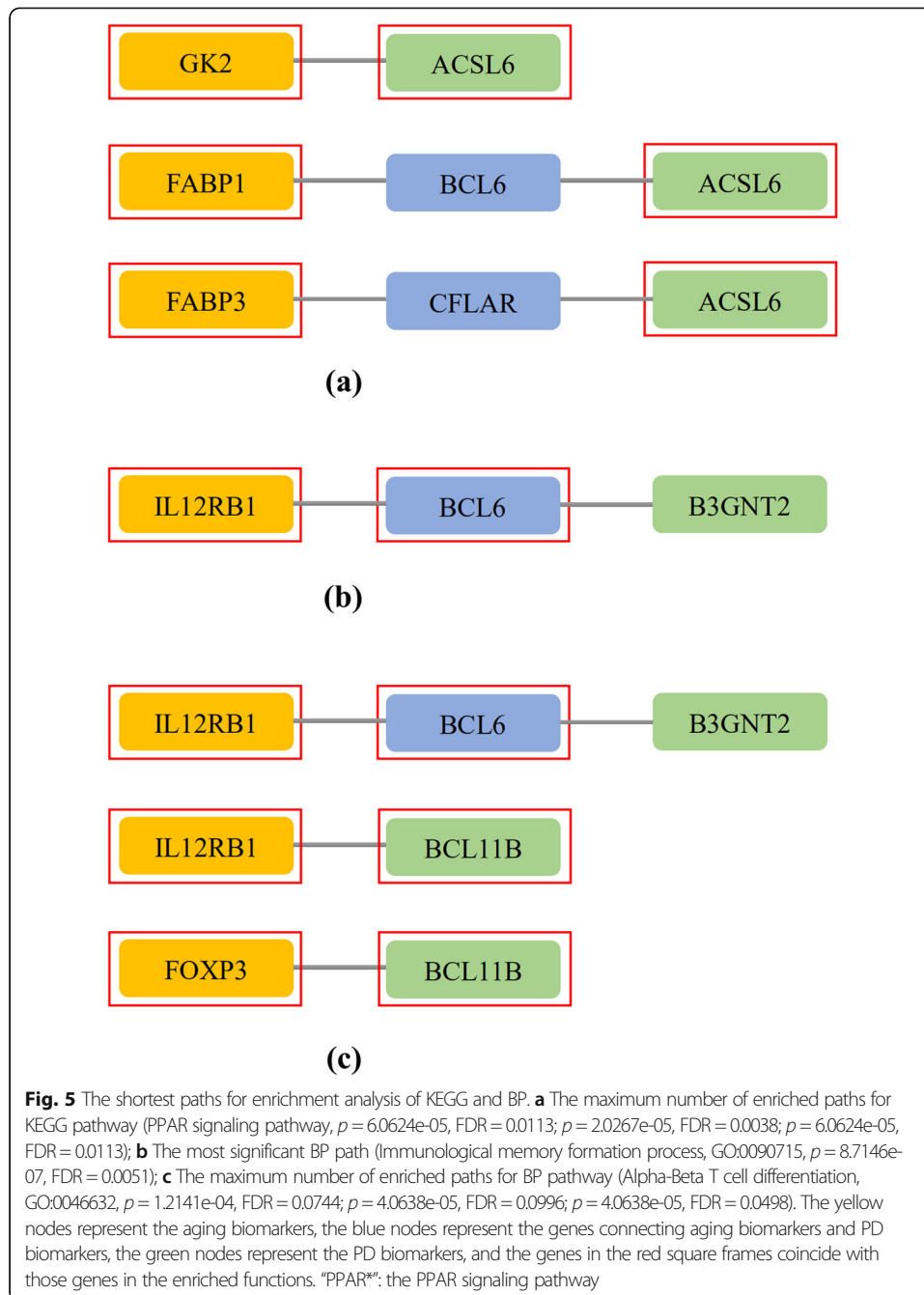
The immunological memory formation process (GO:0090715) was the most enriched BP term ($p = 8.7146e-07$, FDR = 0.0051, Fig. 5b). Memory cells were widely associated with aging. The percentage of memory T cells were with an increasing trend with age, indicating the importance of memory cells during the aging process [29]. Besides, the treatment of CD4+ T cells in PD patients has been drawn more attentions [29]. To sum up, the functions of the immunological memory linking aging acceleration and PD were confirmed thereby.

The alpha and beta T cell differentiation (GO:0046632) was the BP term enriched in most shortest "aging-PD" paths (3 paths, Fig. 5c). There were a series of researches reported that the mechanisms of PD were closed related to the acquired immune system [30], where CD8+ T cells killed SN dopamine neurons through the MHC class I complex [30]; and accumulative α-synuclein peptides (a risk PD factor) could be recognized by T cells [31]. In brief, the T cells contributed to the loss of dopamine neurons.

Table 3 Enrichment analysis results with the most numerous enriched paths (the minimum p-value and FDR)

Name	*P*-value	FDR
KEGG_PPAR_SIGNALING_PATHWAY (3 paths)	6.06e-05	0.0113
	2.03e-05	0.0038
	6.0624e-05	0.0113
GO_ALPHA_BETA_T_CELL_DIFFERENTIATION (3 paths, GO:0046632)	1.21e-04	0.0744
	4.06e-05	0.0996
	4.06e-05	0.0498
GO_POSITIVE_T_CELL_SELECTION (2 paths, GO:0043368)	5.4793e-06	0.0395
	5.4793e-06	0.0298
GO_T_CELL_SELECTION (2 paths, GO: 0045058)	1.0741e-05	0.0395
	1.0741e-05	0.0298

Fig. 5 The shortest paths for enrichment analysis of KEGG and BP. **a** The maximum number of enriched paths for KEGG pathway (PPAR signaling pathway, $p = 6.0624e-05$, FDR = 0.0113; $p = 2.0267e-05$, FDR = 0.0038; $p = 6.0624e-05$, FDR = 0.0113); **b** The most significant BP path (Immunological memory formation process, GO:0090715, $p = 8.7146e-07$, FDR = 0.0051); **c** The maximum number of enriched paths for BP pathway (Alpha-Beta T cell differentiation, GO:0046632, $p = 1.2141e-04$, FDR = 0.0744; $p = 4.0638e-05$, FDR = 0.0996; $p = 4.0638e-05$, FDR = 0.0498). The yellow nodes represent the aging biomarkers, the blue nodes represent the genes connecting aging biomarkers and PD biomarkers, the green nodes represent the PD biomarkers, and the genes in the red square frames coincide with those genes in the enriched functions. "PPAR*": the PPAR signaling pathway

The relationship between aging and PD were reflected by the aging-PD bipartite graphs

To study the relationships between the aging and PD markers, the coefficients were calculated based on each "aging-PD" pair. Then the results of PD and control samples were compared (as mentioned in Materials and Methods). The pair with the highest absolute differential correlation coefficient was GK2 and BCL11B. Evidences reported that GK2 was related to the mitochondrial LC-fatty acid beta-oxidation, where the efficient energy was produced. Energy metabolism was vital to life activities, and was dysregulated during accelerated aging thus in the context of neurodegenerative diseases [32, 33]. BCL11B was a fundamental transcriptional regulator

in cell apoptosis, proliferation, and differentiation [34], thus related to the variation and development of neuronal subtypes of central nervous system [35]. Moreover, it was essential for the development of the immune system [34] and was considered as an interacting partner of T cells [36]. For example, the expression of cytokines (such as IL2, IL5 and IL13) in various T cells was depended on BCL11B [35, 36]. Therefore, BCL11B may be associated with aging and PD through the immune system.

The comprehending correlation mechanisms of aging and PD were further analyzed. As a result, GJC1 was identified the aging marker linking most PD markers and BCL11B was the PD marker linking to most aging markers (Figure S2). Strikingly, the importance of BCL11B was further highlighted. In addition, it has been reported that GJC1 was related to the ion channel activity. Therefore, the dysregulation of GJC1 during advanced aging was associated with ion imbanlances and the neuronal apoptosis [37], where the oxidation of ion channels was also a hallmark of cell deaths in PD progression [37]. In a word, the index of ion imbanlances was a bridge linking aging and PD.

In addtion, a sub-network composed of all the "aging-PD" shortest paths were picked out thereby, and these genes were ranked by their betweennesses (in descending order) along with the permutation *p*-values (shown in Table 4). There were 4 out of the top 10 genes whose p-values were significant. For example, BCL6 was with the largest betweenness 80 (the permutation p-value = 0), meaning that there were 80 shortest paths going through this gene. Obviously, BCL6 is a critical marker in cell apoptosis. It has also been reported that BCL6 contributed to the differentiation of follicular helper T cells and the inflammation activation of macrophages [38]. Therefore, it could be speculated that BCL6 was one of the most important markers linking accelerated aging and LOPD.

Discussion

In this work, the relationship between aging and LOPD were studied through our computational pipeline. First, the aging markers were identified by the aging predictor. Thus, the improved PD predictor compared with the normal aging process was modeled to identify PD markers. Besides, the aging scores showed the aging acceleration pattern in LOPD compared with the control group. Further, the aging acceleration network based on aging scores was constructed, where the aging-PD paths revealed critical functions at system level.

Table 4 The top 10 genes with the top betweenness in the aging acceleration network

Gene	Betweenness	*p*-value
BCL6	80	0*
ARPC5L	69	0.076
APBB1IP	39	0*
ASNS	23	0.081
BTN2A1	20	0.014*
CFLAR	15	0.003*
ANK3	11	0.082
CD2AP	11	0.195
ABHD3	9	0.340
CASP2	9	0.354

*:*p*-value< 0.05, significant

The top aging marker (RBMS1) played a vital role in cell apoptosis [17]; the most notable PD risk biomarker (ADD2) revealed the importance of the mitochondrial Ca^{2+} imbalance; and the most important network node SLC37A1 also indicated the mitochondria-related mechanisms of PD through the energy metabolism. Four out of top ten genes with the significant betweennesses suggested that the immune system was vulnerable to be dysregulated during advanced aging [39]. More importantly, the enrichment results highlighted the crucial core of the PPAR signaling pathway interacted with the neuroinflammation, mitochondrial oxidative stress and cell apoptosis. The most closely related "aging-PD" pair was GK2 and BCL11B, indicating the critical role of GK2 that was related to ATP production in mitochondria. BCL11B was also as the PD biomarker linked most aging biomarkers and associated to the immune response and eliminate inflammation. What's more, GJC1 (as the aging marker linked the most PD markers) acted as a key role in the ion channel. In short, our results integrated aging and PD markers systematically (Fig. 6).

In the neuroimmune-endocrine theories of aging, the neuroinflammation response was exaggerated and prolonged in the healthy aged brain, thus advanced aging tended to induce a series of inflammatory factors amplified this response [40]. Our results further confirmed the importance of inflammation in the progression of PD [41]. In addition, apart from the mechanism of neuroinflammation-induced PD [42], other mechanisms of LOPD were integrated through our results. More interestingly, we proposed that the improved PD predictor was more reliable in exploring how PD was induced by aging acceleration. The PD markers identified by our improved prediction model have confirmed the dysfunctions in the mitochondrial Ca2+ pathway [43]. Further, the neurodegenerative disease was more vulnerable to be induced by excessive apoptosis in neuronal cells [44], thus the network markers with significant high betweenness also indicated the vital role of cell apoptosis, where the ion imbalances also played an important role in PD progression (i.e. the function of GJC1). Strikingly, the enrichment results demonstrated that the PPAR signaling pathway played a central role in regulating the inflammation response, energy metabolisms and the cell apoptosis. In summary, our results integrated the neuronal apoptosis, neuroinflammations, mitochondrial metabolisms, and the ion imbalances as comprehensive mechanisms of PD, where the PPAR signaling pathway may play a central role on the basis of immune disorders (Fig. 6).
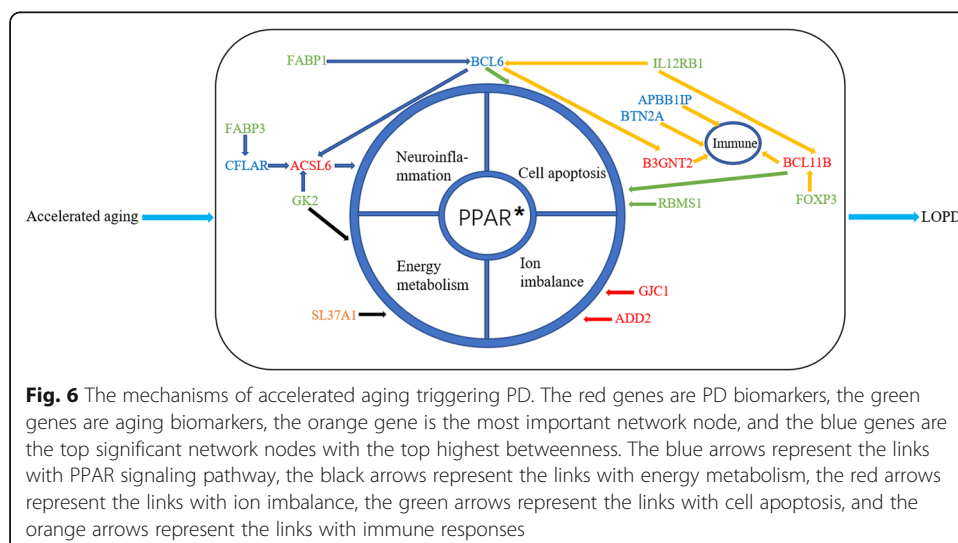
Fig. 6 The mechanisms of accelerated aging triggering PD. The red genes are PD biomarkers, the green genes are aging biomarkers, the orange gene is the most important network node, and the blue genes are the top significant network nodes with the top highest betweenness. The blue arrows represent the links with PPAR signaling pathway, the black arrows represent the links with energy metabolism, the red arrows represent the links with ion imbalance, the green arrows represent the links with cell apoptosis, and the orange arrows represent the links with immune responses

In brief, our results indicated that LOPD was induced by accelerated aging, along with dysfunctions in the immune system [45]. In addition, the neuronal apoptosis, mitochondrial metabolic processes, the neuroinflammation and the ion imbanlances interacted with each other to induce PD coordinately. It has also been reported the homeostasis of these mechanisms would be disrupted during the brain aging, presenting the key risk factor for neurodegeneration [46]. Therefore, our results brought insight into a comprehensive understanding of LOPD development triggered accelerated aging.

Conclusion

To conclude, the aging predictor and the improved PD predictor revealed potential risk biomarkers by machine learning methods, respectively. Then the aging scores were calculated to assess the aging process in both PD and control groups. The aging scores revealed significant aging acceleration in PD. Further, the aging acceleration network was constructed by comparing (partial) correlations of genes between LOPD and control samples. Additionally, the Aging-PD bipartite graph was utilized to investigate the relationship between aging and LOPD. As a result, the enrichment analysis indicated the critical mechanisms of LOPD during advanced aging. That is to say, during accelerated aging, the cell apoptosis, mitochondrial disorders, the ion imbanlances and the increased neuroinflammation combined to trigger PD coordinately, where the PPAR signaling pathway probably played the most critical role in immune dysfunctions.

Materials and methods

Gene expression profiles and data preprocessing

The gene expression profiles were obtained in Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/), including GSE8397, GSE20295, GSE28894, GSE57475, GSE99039, and GSE15745 (Text S2). All of the selected PD or normal samples were not only with accurate age information, but also with more than 10 sample sizes. These samples were obtained from five different platforms: GPL96, GPL97, GPL570, GPL6104, and GPL6947.

The steps of obtaining gene expression profiles were as follows:

(1) The probes of samples were converted to various gene symbols of platforms.
(2) Gene expression values with the same gene symbols were summarized.
(3) The total data matrix was integrated and the missing gene expression values were filled with 0.
(4) The genes with zero expression values ≥30% were deleted.
(5) Patients with early-onset PD (age < 50) were removed.

Finally, we collected the expression datasets of 425 healthy young samples (285 in training data and 140 in test data), 447 normal aged samples (300 + 147) and 392 PD samples (265 + 127) including 13,883 gene features (Text S1-S2).

In addition, the profiles in each platform was further normalized, seperately:

(1) On account of the mean and the standard deviation of the normal samples, the z-score normalization was performed for both normal old samples and PD samples.

(2) The Singular Value Decomposition (SVD) method was performed to eliminate the inter-sample variation based on the top three principal components of the normal and PD samples, respectively.

(3) The z-score was then utilized to normalize all samples based on the mean and the standard deviation of the normal old samples.

Modeling the aging predictor and identifying aging biomarkers

To model the reliable aging predictor, the normal samples were divided into training dataset and test dataset in which the ratio of training dataset samples to test dataset samples were close to 2:1. In addition, the number of young samples (< 50 years old) was similar to the number of old samples (≥50 years old) in both the training dataset and the test dataset. To evaluate the gene-age correlation, the Pearson correlation coefficient between genes and age labels was calculated:

$$corr(gene, AgeGroup) \tag{1}$$

Then, 13,883 genes were sorted in descending order based on the absolute values of the correlation coefficients. Based on sorted gene symbols, k-NN (k = 5 with cosine distance) was performed to construct the top 100 aging predictors. The optimal model would be selected through the 10-fold cross-validation. To prove the predictability and efficiency of the aging predictor, the selected model was validated in the test dataset.

Modeling the improved PD predictor and identifying PD biomarkers

To uncover the mechanism of LOPD in the aging acceleration process, normal young samples were considered as the auxiliary outlier group, and the improved PD predictor was modeled by comparing LOPD and normal aged samples (age ≥ 50). The deviations of PD and normal aged samples relative to healthy young samples were calculated as follows:

(1) The first 3 principal components (row vectors) of healthy young samples (row vectors were samples, column vectors were genes) were extracted;

(2) The original gene expression profiles of the normal aged and PD samples were substituted based on the residuals by regressing these 3 principal components of the healthy young samples, respectively.

Further, the new gene datasets from the normal aged and PD samples were divided into training dataset and test dataset. Similarly, the ratio of the training samples and the test samples was closed to 2:1, and the proportion of the normal old samples and PD samples was approximate in the training data and test data, respectively. The PD aging predictor was also constructed by sorting descending absolute values of Pearson correlation coefficients and utilizing k-NN (k = 5 with the cosine distance) method. The traditional PD predictor using original gene datasets was also calculated to compare and evaluate the prediction ability of the improved LOPD model.

Calculating the aging scores

For all normal samples and PD samples, the quantitative aging scores were further summarized based on chronological age using the aging markers identified by kNN. The detailed pipeline was as follows:

(1) The chronological age was transformed using the sigmoid function:

$$transformated\ age = \frac{1}{1 + \exp(-(age-50)/50)} \tag{2}$$

(2) The transformated age was predicted based on the aging markers using the linear regression:

$$transformated\ age = \sum b_i * aging\_mar\ker_i \tag{3}$$

where b_i was the regression coefficient for each aging marker, respectively.

(3) The prediction of transformated age was used as the aging score;
(4) The Kruskal-Wallis test was used to compare the accerelated aging pattern between LOPD and the normal aged samples for different age groups.

The construction of the aging acceleration network

The aging acceleration network based on the aging scores was constructed to further reveal the relationship between aging and PD. For the purpose of reliable network validation, the aging accelerated networks were constructed on the basis of training dataset and test dataset, respectively. First of all, for the respective 13,883 gene values of normal old samples and PD samples, the Pearson correlation coefficient as well as the partial correlation coefficient (based on the aging score) of any each pair of genes was calculated, respectively. Further, the assessment criteria of statistical significance was the Benjamini-Hochberg False Discovery Rates (FDR) < 0.05. Additionally, the absolute difference of partial correlation coefficient between PD samples and normal aged samples was calculated, as well as the absolute difference of correlation coefficient. If both of the absolute differences were greater than 0.5, the relationship between pairs of genes was further retained. Finally, with Fisher's exact test, the similarity of two aging acceleration networks based on training dataset and test dataset would be tested. At the same time, scale-free property of the aging acceleration network was validated.

The construction of the aging-PD bipartite graph

To better understand the detailed connections between aging biomarkers and PD biomarkers, the Aging-PD bipartite graphs were constructed in the context of the aging acceleration network, depending on the existed shortest paths (using the Dijkstra algorithm) from each aging marker to PD marker. To find out each pair of notable aging markers and PD markers, the absolute value of the difference between two correlation coefficients in the normal aged group and PD group was calculated, and the "aging-PD" pair with the opposite sign was further retained. Then the Aging-PD biomarker pair with the maximum absolute value should be noted. Moreover, the aging biomarker that most closely linked to PD markers and the PD biomarker that most closely linked to aging biomarkers were important. Therefore, for each aging marker, the sum of absolute values of PD markers was calculated. For each PD marker, the sum of absolute values of aging markers was also calculated. As a result, the aging/PD marker with high accumulated absolute differential correlations were the considered as critical genes

interacted with PD/aging markers. As a result, totally 362 "aging-PD" pairs were identified out of 552 (69*8) pairs.

Further, the subnetwork with shortest pathways among selected "aging-PD" pairs was constructed, and genes in the subnetwork were sorted by their betweennesses in descending order. To test whether the top betweenness genes were hubs in the background network or not, we ran a permutation to count the occurrence times of the top genes in the shortest paths between random selected genes (containing the same numbers of "aging-PD" pairs, 69*8) when they were with greater betweennesses than those in our study. We repeated this process 1000 times, and the *p*-value was calculated as the proportion of occurrence times of the top betweenness genes in 1000 permutations.

Enrichment analysis

To find vital biological functions, enrichment analysis was utilized. Gene Ontology (GO) terms and KEGG pathways were downloaded from Gene Set Enrichment Analysis (GSEA) platform (http://software.broadinstitute.org/gsea/downloads.jsp, version 7.0). The hypergeometric test was used to performed to estimate the enrichment of KEGG pathways or GO BP terms. The hypergeometric test formula was given as

$$P(X \geq x) = 1 - \sum_{k=0}^{x-1} \frac{C_M^k * C_{N-M}^{n-k}}{C_N^k} \qquad (4)$$

where N is the total gene number of the gene sets, M shows the number of known genes (i.e. KEGG pathway, GO terms), n is the number of identified genes in each shortestpath, and k is the number of common genes between the known genes and the identified candidate genes (in each aging-PD shortest path). The p-value of each pathway was controlled by the Benjamini-Hochberg method. The final thresholds were $p < 0.05$ and FDR < 0.1.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s13040-020-00215-w.

Additional file 1 : **Figure S1.** The learning curve and ROC curve in the traditional PD predictor. **Figure S2.** The heatmap about the differential correlation coeffecients between the aging biomarkers and PD biomarkers. Xlable reprensents PD markers, Ylable represents aging markers.

Additional file 2 : **Text S1.** The gene set used in this work.

Additional file 3 : **Text S2.** The detailed datasets in this work.

Additional file 4 : **Table S1.** The top 69 aging biomarkers and their correlation coefficients. **Table S2.** The top 8 PD markers in the improved PD predictor and their correlation coefficients. **Table S3.** The top 94 PD markers in the traditional PD predictor and their correlation coefficients.

Availability of data and materials
The data supporting the results of this article are included and cited within the article and its additional files (i.e Text S1, Table S1 and Text S2).

Ethics approval and consent to participate
Because the human data of our manuscript is all from the GEO platform in NCBI, we believe that we do not need additional ethics approval.

Consent for publication
Not applicable.

Competing interests
The authors declared that they have no competing interests.

Author details
[1]Department of Biomedical Engineering, School of Fundamental Sciences, China Medical University, Shenyang 110122, Liaoning Province, China. [2]China Medical University, The Queen's University of Belfast Joint College, China Medical University, Shenyang 110122, Liaoning Province, China. [3]Bio-Med Big Data Center, Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institute of Nutrition and Health, Shanghai Institutes for Biological Sciences, University of Chinese Academy of Sciences, Chinese Academy of Science, Shanghai 200031, China. [4]School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China. [5]Collaborative Innovation Center of Genetics and Development, Fudan University, Shanghai 200433, China. [6]Shanghai Center for Bioinformation Technology, Shanghai Academy of Science and Technology, Shanghai 201203, China. [7]Tumor Etiology and Screening Department of Cancer Institute and General Surgery, The First Affiliated Hospital of China Medical University, 155# North Nanjing Street, Heping District, Shenyang City 110001, Liaoning Province, China.

References
1. Xicoy H, Peñuelas N, Vila M, Laguna A. Autophagic- and lysosomal-related biomarkers for Parkinson's disease: lights and shadows. Cells. 2019;8(11):1317.
2. Delenclos M, Jones DR, McLean PJ, Uitti RJ. Biomarkers in Parkinson's disease: advances and strategies. Parkinsonism Relat Disord. 2016;22:S106–10.
3. Collier TJ, Kanaan NM, Kordower JH. Aging and Parkinson's disease: different sides of the same coin? Mov Disord. 2017; 32(7):983–90.
4. Frisardi V, Santamato A, Cheeran B. Parkinson's disease: new insights into pathophysiology and rehabilitative approaches. Parkinsons Dis. 2016;2016:3121727.
5. Bobela W, Aebischer P, Schneider BL. Alpha-synuclein as a mediator in the interplay between aging and Parkinson's disease. Biomolecules. 2015;5(4):2675–700.
6. Liu J, Liu W, Lu Y, Tian H, Duan C, Lu L, Gao G, Wu X, Wang X, Yang H. Piperlongumine restores the balance of autophagy and apoptosis by increasing BCL2 phosphorylation in rotenone-induced Parkinson disease models. Autophagy. 2018;14(5):845–61.
7. Argüelles S, Guerrero-Castilla A, Cano M, Muñoz MF, Ayala A. Advantages and disadvantages of apoptosis in the aging process. Ann N Y Acad Sci. 2019;1443(1):20–33.
8. Calabrese V, Santoro A, Monti D, Crupi R, Di Paola R, Latteri S, Cuzzocrea S, Zappia M, Giordano J, Calabrese EJ, Franceschi C. Aging and Parkinson's disease: Inflammaging, neuroinflammation and biological remodeling as key factors in pathogenesis. Free Radic Biol Med. 2018;115:80–91.
9. Rudow G, O'Brien R, Savonenko AV, Resnick SM, Zonderman AB, Pletnikova O, Marsh L, Dawson TM, Crain BJ, West MJ, Troncoso JC. Morphometry of the human substantia nigra in ageing and Parkinson's disease. Acta Neuropathol. 2008; 115(4):461–70.
10. Chesnokova AY, Ekimova IV, Pastukhov YF. Parkinson's disease and aging. Adv Gerontol. 2018;31(5):668–78.
11. Lahmiri S, Dawson DA, Shmuel A. Performance of machine learning methods in diagnosing Parkinson's disease based on dysphonia measures. Biomed Eng Lett. 2018;8(1):29–39.
12. Latourelle JC, Beste MT, Hadzi TC, Miller RE, Oppenheim JN, Valko MP, Wuest DM, Church BW, Khalil IG, Hayete B, Venuto CS. Large-scale identification of clinical and genetic predictors of Parkinson's disease motor progression in newly-diagnosed patients: a longitudinal cohort study and validation. Lancet Neurol. 2017;16(11):908–16.
13. Lahmiri S, Shmuel A. Detection of Parkinson's disease based on voice patterns ranking and optimized support vector machine. Biomed Signal Process Control. 2019;49:427–33.
14. Lahmiri S. Gait nonlinear patterns related to Parkinson's disease and age. IEEE Trans Instrum Meas. 2018;68(7):2545–51.
15. Sood S, Gallagher IJ, Lunnon K, Rullman E, Keohane A, Crossland H, Phillips BE, Cederholm T, Jensen T, van Loon LJC, Lannfelt L, Kraus WE, Atherton PJ, Howard R, Gustafsson T, Hodges A, Timmons JA. A novel multi-tissue rna diagnostic of healthy ageing relates to cognitive health status. Genome Biol. 2015;16(1):185.
16. Hubberten M, Bochenek G, Chen H, Häsler R, Wiehe R, Rosenstiel P, Jepsen S, Dommisch H, Schaefer AS. Linear isoforms of the long noncoding RNA CDKN2B-AS1 regulate the c-myc-enhancer binding factor RBMS1. Eur J Hum Genet. 2019; 27(1):80–9.
17. Yin M, Lü M, Yao G, Tian H, Lian J, Liu L, Liang M, Wang Y, Sun F. Transactivation of microRNA-383 by steroidogenic factor-1 promotes estradiol release from mouse ovarian granulosa cells by targeting RBMS1. Mol Endocrinol. 2012;26(7): 1129–43.
18. Kirkland JL, Tchkonia T. Cellular senescence: a translational perspective. EBioMedicine. 2017;21:21–8.

Li *et al. BioData Mining*   (2020) 13:4

Page 17 of 17

19. Liu CM, Hsu WH, Lin WY, Chen HC. Adducin family proteins possess different nuclear export potentials. J Biomed Sci. 2017;24(1):30.
20. Kiang KM, Leung GK. A review on adducin from functional to pathological mechanisms: future direction in cancer. Biomed Res Int. 2018;2018:3465929.
21. Ludtmann MHR, Abramov AY. Mitochondrial calcium imbalance in Parkinson's disease. Neurosci Lett. 2018;663:86–90.
22. Cappello AR, Curcio R, Lappano R, Maggiolini M, Dolce V. The physiopathological role of the exchangers belonging to the SLC37 family. Front Chem. 2018;6:122.
23. Nakano M, Imamura H, Sasaoka N, Yamamoto M, Uemura N, Shudo T, Fuchigami T, Takahashi R, Kakizuka A. ATP maintenance via two types of ATP regulators mitigates pathological phenotypes in mouse models of Parkinson's disease. EBioMedicine. 2017;22:225–41.
24. Zhao L, Zheng X, Liu J, Zheng R, Yang R, Wang Y, Sun L. PPAR signaling pathway in the first trimester placenta from *in vitro* fertilization and embryo transfer. Biomed Pharmacother. 2019;118:109251.
25. Burgaz S, García C, Gómez-Cañas M, Muñoz E, Fernández-Ruiz J. Development of an oral treatment with the PPAR-γ-acting cannabinoid VCE-003.2 against the inflammation-driven neuronal deterioration in experimental Parkinson's disease. Molecules. 2019;24(15):2702.
26. Machado MMF, Bassani TB, Cóppola-Segovia V, Moura ELR, Zanata SM, Andreatini R, Vital MABF. PPAR-γ agonist pioglitazone reduces microglial proliferation and NF-κB activation in the Substantia Nigra in the 6-hydroxydopamine model of Parkinson's disease. Pharmacol Rep. 2019;71(4):556–64.
27. Chaturvedi RK, Beal MF. PPAR-γ: therapeutic prospects in Parkinson's disease. J Neurochem. 2008;106(2):506–18.
28. Corona JC, Duchen MR. PPARγ and PGC-1α as therapeutic targets in Parkinson's. Neurochem Res. 2015;40(2):308–16.
29. Fiszer U, Mix E, Fredrikson S, Kostulas V, Link H. Parkinson's disease and immunological abnormalities: increase of HLA-DR expression on monocytes in cerebrospinal fluid and of CD45RO+ T cells in peripheral blood. Acta Neurol Scand. 1994;90(3):160–6.
30. Sulzer D, Alcalay RN, Garretti F, Cote L, Kanter E, Agin-Liebes J, Liong C, McMurtrey C, Hildebrand WH, Mao X, Dawson VL, Dawson TM, Oseroff C, Pham J, Sidney J, Dillon MB, Carpenter C, Weiskopf D, Phillips E, Mallal S, Peters B, Frazier A, Lindestam Arlehamn CS, Sette A. T cells of Parkinson's disease patients recognize α–synuclein peptides. Nature. 2017; 546(7660):656–61.
31. Fyfe I. Parkinson disease: T cells recognize α-Synuclein peptides in Parkinson disease. Nat Rev Neurol. 2017;13(8):450–1.
32. Błaszczyk JW. The emerging role of energy metabolism and neuroprotective strategies in Parkinson's disease. Front Aging Neurosci. 2018;10:301.
33. Camandola S, Mattson MP. Brain metabolism in health, aging, and neurodegeneration. EMBO J. 2017;36(11):1474–92.
34. Lessel D, Gehbauer C, Bramswig NC, Schluth-Bolard C, Venkataramanappa S, van Gassen KLI, Hempel M, Haack TB, Baresic A, Genetti CA, Funari MFA, Lessel I, Kuhlmann L, Simon R, Liu P, Denecke J, Kuechler A, de Kruijff I, Shoukier M, Lek M, Mullen T, Lüdecke HJ, Lerario AM, Kobbe R, Krieger T, Demeer B, Lebrun M, Keren B, Nava C, Buratti J, Afenjar A, Shinawi M, Guillen Sacoto MJ, Gauthier J, Hamdan FF, Laberge AM, Campeau PM, Louie RJ, Cathey SS, Prinz I, Jorge AAL, Terhal PA, Lenhard B, Wieczorek D, Strom TM, Agrawal PB, Britsch S, Tolosa E, Kubisch C. BCL11B mutations in patients affected by a neurodevelopmental disorder with reduced type 2 innate lymphoid cells. Brain. 2018;141(8):2299–311.
35. Fang D, Cui K, Hu G, Gurram RK, Zhong C, Oler AJ, Yagi R, Zhao M, Sharma S, Liu P, Sun B, Zhao K, Zhu J. Bcl11b, a novel GATA3-interacting protein, suppresses Th1 while limiting Th2 cell differentiation. J Exp Med. 2018;215(5):1449–62.
36. Lennon MJ, Jones SP, Lovelace MD, Guillemin GJ, Brew BJ. Bcl11b-a critical neurodevelopmental transcription factor-roles in health and disease. Front Cell Neurosci. 2017;11:89.
37. Patel R, Sesti F. Oxidation of ion channels in the aging nervous system. Brain Res. 2016;1639:174–85.
38. Kutyavin VI, Chawla A. BCL6 regulates brown adipocyte dormancy to maintain thermogenic reserve and fitness. Proc Natl Acad Sci U S A. 2019;116(34):17071–80.
39. Pinti M, Appay V, Campisi J, Frasca D, Fülöp T, Sauce D, Larbi A, Weinberger B, Cossarizza A. Aging of the immune system: focus on inflammation and vaccination. Eur J Immunol. 2016;46(10):2286–301.
40. Barrientos RM, Frank MG, Watkins LR, Maier SF. Aging-related changes in neuroimmune-endocrine function: implications for hippocampal-dependent cognition. Horm Behav. 2012;62(3):219–27.
41. Dzamko N, Geczy CL, Halliday GM. Inflammation is genetically implicated in Parkinson's disease. Neuroscience. 2015;302:89–102.
42. Kempuraj D, Thangavel R, Natteru PA, Selvakumar GP, Saeed D, Zahoor H, Zaheer S, Iyer SS, Zaheer A. Neuroinflammation Induces Neurodegeneration. J Neurol Neurosurg Spine. 2016;1(1):1003.
43. Duda J, Pötschke C, Liss B. Converging roles of ion channels, calcium, metabolic stress, and activity pattern of Substantia nigra dopaminergic neurons in health and Parkinson's disease. J Neurochem. 2016;139(Suppl 1):156–78.
44. Higami Y, Shimokawa I. Apoptosis in the aging process. Cell Tissue Res. 2000;301(1):125–32.
45. Liang Z, Zhao Y, Ruan L, Zhu L, Jin K, Zhuge Q, Su DM, Zhao Y. Impact of aging immune system on neurodegeneration and potential immunotherapies. Prog Neurobiol. 2017;157:2–28.
46. Pluvinage JV, Wyss-Coray T. Systemic factors as mediators of brain homeostasis, ageing and neurodegeneration. Nat Rev Neurosci. 2020;21(2):93–102.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.