REVIEW

Translational Informatics for Parkinson's Disease: from Big Biomedical Data to Small Actionable Alterations

Bairong Shen [1,*,a], Yuxin Lin [2,b], Cheng Bi [2,c], Shengrong Zhou [2,d]
Zhongchen Bai [3,e], Guangmin Zheng [3,f], Jing Zhou [3,g]

[1] *Institutes for Systems Genetics, West China Hospital, Sichuan University, Chengdu 610041, China*
[2] *Center for Systems Biology, Soochow University, Suzhou 215006, China*
[3] *Center for Translational Biomedical Informatics, Guizhou University School of Medicine, Guiyang 550025, China*

Abstract **Parkinson's disease** (PD) is a common neurological disease in elderly people, and its morbidity and mortality are increasing with the advent of global ageing. The traditional paradigm of moving from small data to big data in biomedical research is shifting toward big data-based identification of small actionable alterations. To highlight the use of big data for precision PD medicine, we review PD big data and informatics for the translation of basic PD research to clinical applications. We emphasize some key findings in clinically actionable changes, such as susceptibility genetic variations for PD risk population screening, biomarkers for the diagnosis and stratification of PD patients, risk factors for PD, and lifestyles for the prevention of PD. The challenges associated with the collection, storage, and modelling of diverse big data for PD precision medicine and **healthcare** are also summarized. Future perspectives on **systems modelling** and intelligent medicine for PD monitoring, diagnosis, treatment, and healthcare are discussed in the end.

Introduction

The disease spectrum is changing with the ballooning of elderly society. The morbidity and mortality of geriatric disease, including Alzheimer's disease (AD) and Parkinson's disease (PD), are increasing globally [1]. The social burden of the care of elderly patients is becoming a considerable challenge because of the lack of sufficient medical and labour resources. The shortage of medical care resources and the increasing

* Corresponding author.
  E-mail: bairong.shen@scu.edu.cn (Shen B).
[a] ORCID: 0000-0003-2899-1531.
[b] ORCID: 0000-0001-6543-0317.
[c] ORCID: 0000-0002-8076-2704.
[d] ORCID: 0000-0002-0190-877X.
[e] ORCID: 0000-0002-0286-4246.
[f] ORCID: 0000-0002-3781-170X.
[g] ORCID: 0000-0002-4259-1158.

demand of the ageing society will be obstacles to social and economic development.

PD is one of the most common neurodegenerative diseases (NDDs) in elderly people. As the most frequent movement disorder, PD usually develops very slowly, although it can be accelerated in the latter years. It can take more than 20 years to proceed the beginning of neurodegeneration to the appearance of prodromal symptoms *en route* and to the manifestation of typical clinical symptoms of PD [2]. A search of the PubMed database with the term "Parkinson's disease [tiab] OR Parkinson disease [tiab]" retrieves more than 87,500 records of PD studies at present. Nonetheless, the causative and molecular mechanism of PD remains elusive, although it is generally believed to involve complex interactions between genetics [3], gut microbiota [4], environmental factors [5], as well as unhealthy lifestyles [6]. These complex interactions pose great challenges in gaining a comprehensive understanding of the holistic mechanism underlying PD pathogenesis and progression.

Early diagnosis and prevention of PD is preferred over late clinical treatment of the disease because it can alleviate both social demand and family burden. Many basic questions remain to be addressed for PD studies before potential translation, such as the identification of biomarkers for personalized diagnosis and stratification of patients [7], the discovery of genetic or environmental factors for the screening of highly susceptible populations, and the finding of a positive lifestyle to facilitate personalized healthcare of elderly people [8,9]. To investigate the molecular mechanisms underlying PD and answer the aforementioned questions, sufficient data and information about the genotypes and clinical phenotypes of different subtypes of PD are prerequisite to model the complex interactions.

In recent decades, we have witnessed a rapid development of biotechnologies, especially high-throughput sequencing technologies. Deep sequencing for genetic architecture, gene expression and epigenetic patterns is becoming less expensive, and the costs of whole-genome sequencing have decreased from hundreds of million dollars to hundreds of dollars. Not only has the sequencing data accumulated at an unprecedented rate, the physiological data collected from different wearable sensors, the biochemical data detected by point-of-care tests, and the medical imaging data, including magnetic resonance imaging (MRI) and positron emission tomography computed tomography (PET-CT), are also increasing rapidly. We are now in a big data and digital medicine era [10,11]. Data from healthy people as well as preclinical and clinical data from patients together contribute to the big volume of big PD data for future data-driven medicine (Figure 1).

The 5 Vs of big data characterization for PD are shown in Figure 1. The data formats vary and include strings for genetic data, images for PD brain structures, unstructured or semi-structured formats with real values, text for electronic health record (EHR), and time series data for physiological signals. The digitalization of diverse measurements speeds up the generation of all kinds of PD data. In particular, wearable sensors combined with smart phones make it possible to collect data real-time and obtain dynamic electroencephalograph (EEG). These would help with the monitoring and diagnosis of PD patients [12]. The identification of actionable key players and alterations from the considerably large, noisy and diverse

unstructured big data is the goal for translational informatics studies. In this review, we mainly discuss the value of PD big data mining, as well as the challenges and perspectives for the translation of PD big data to valuable biomarker discovery and risk factor screening for the future clinical management and healthcare of PD.

Actionable alterations for PD diagnosis and prevention

The traditional paradigm of translational research for disease biomarker or risk factor discovery is often from small data to big data. It starts from a hypothesis-driven investigation of the biological functions of few genes, proteins, or other biological molecules, followed by test of their biological functions and medical roles, moving from cell lines, animal models, and a small number of patients to big population validation. Biomarker and/or drug discoveries often fail in last-phase trials because the features or discoveries obtained from small data do not always work well in a big and diverse data space.

Nowadays, the paradigm of biomedical research is shifting to one involving a move from big data to small data. Identifying small but important actionable alterations from big data mining and systems biological modelling is becoming possible. In this section, we discuss the clinically actionable alterations from four aspects. These include PD susceptibility genetic variants, biomarkers for PD diagnosis and prognosis, non-genetic PD risk factors, and lifestyles positively or negatively affecting PD (Figure 2).

Genetic alterations and susceptibility to PD

Before 1997, when mutations in *SNCA*, the gene encoding synuclein alpha, were screened from PD families, PD was recognized as a sporadic and typical non-genetic disorder [13]. With the increasing number of genetic risks identified, PD is now considered a disorder ranging from monogenic to polygenic inheritance associated with a complex interaction between genetics, lifestyle, and environmental exposures [14].

Table 1 provides a partial lists of previously reported genetic risk factors for familial or sporadic PD, covering single-nucleotide polymorphisms (SNPs), haplotypes, copy number variations, and other polymorphisms. In our Neurodegenerative Disease Variation Database (NDDVD), more than 600 variants in 43 genes associated with PD have been collected [14]. Some of the genetic variants could be driver mutations, such as mutations in *SNCA* and *LRRK2*. *LRRK2* is a key player and a common inheritable factor in PD. It encodes leucine rich repeat kinase 2, a kinase involved in the signalling pathways related to neuronal death, and could be a potential therapeutic target for PD [15–17]. While other variants are mostly passenger but not driver mutations, they can work together to disrupt the biological system and cause PD [18]. At present, both the common disease-multiple rare variant (CDMV) and common disease-common variant (CDCV) hypotheses could be applied to explain the genetic variants for PD [19], and the penetrance of each gene variant may be associated with different populations, ages, genders, ethnicities, *etc*. Most of the complex cases cannot be reasonably explained yet. The cumulative effects of the variants on PD
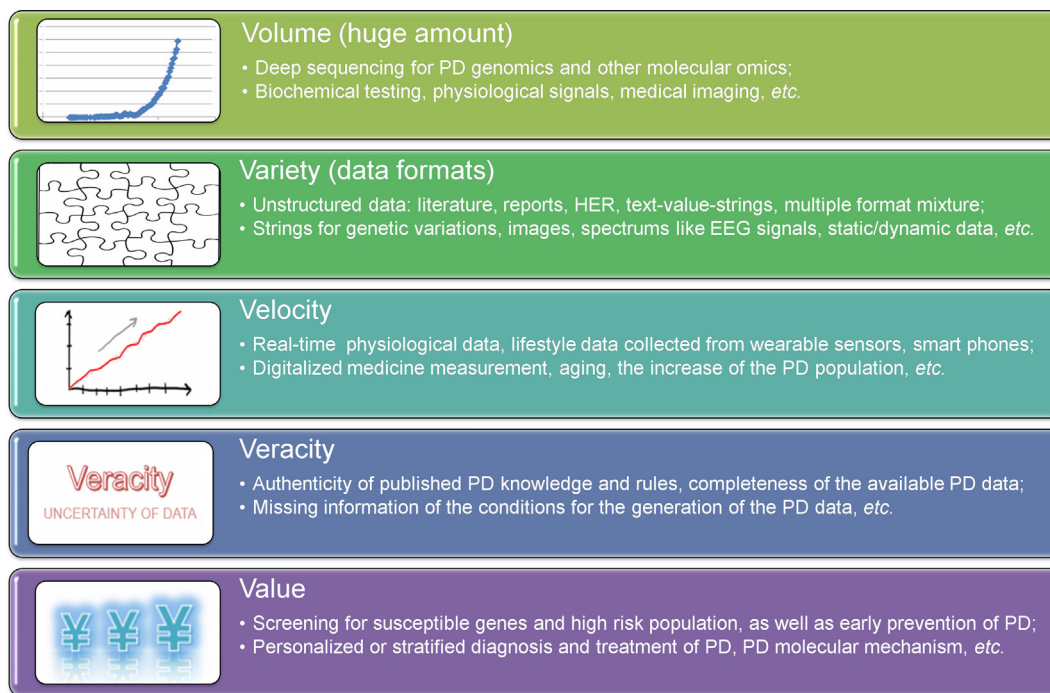
Volume (huge amount)
- Deep sequencing for PD genomics and other molecular omics;
- Biochemical testing, physiological signals, medical imaging, *etc.*

Variety (data formats)
- Unstructured data: literature, reports, HER, text-value-strings, multiple format mixture;
- Strings for genetic variations, images, spectrums like EEG signals, static/dynamic data, *etc.*

Velocity
- Real-time physiological data, lifestyle data collected from wearable sensors, smart phones;
- Digitalized medicine measurement, aging, the increase of the PD population, *etc.*

Veracity
Veracity
UNCERTAINTY OF DATA
- Authenticity of published PD knowledge and rules, completeness of the available PD data;
- Missing information of the conditions for the generation of the PD data, *etc.*

Value
- Screening for susceptible genes and high risk population, as well as early prevention of PD;
- Personalized or stratified diagnosis and treatment of PD, PD molecular mechanism, *etc.*

Figure 1   The 5 Vs of PD big data
PD, Parkinson's disease; EHR, electronic health record; EEG, electroencephalograph.
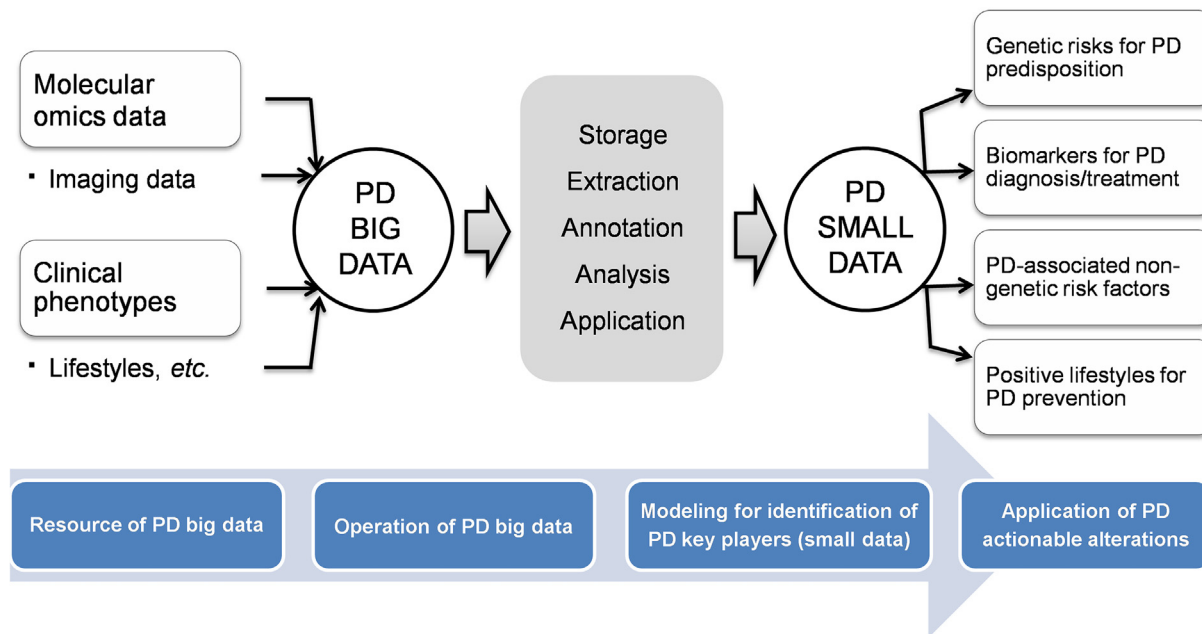

Figure 2   PD translational informatics: from big data to small alterations

need to be evaluated on an individual basis. Two challenges remain for the genetic study of PD. The first is the discovery and curation of more variants, as each of them may have only a small effect on the pathogenesis and progression of PD, and the second is the building of models to accurately predict the cumulative effect of these genetic variants.

PD-associated non-genetic risk factors

Although genetic factors could be important or even act as the driving force behind the pathogenesis and progression of PD, genetic susceptibility can explain only a small portion of PD cases. Many non-genetic factors are found to increase the risk

Table 1   PD genetic risk factors

Note: Variants were mentioned in different formats in previous publications and renamed in this article per the recommendation of Human Genome Variation Society (HGVS) for consistency. PD, Parkinson's disease; SNP, single nucleotide polymorphism; CNV, copy number variation; fs, frame shift; IVS, intervening sequence.

Gene	Variable sites	PMID
LRRK2	SNPs in rs10878226 and rs11176013; p.Gly2385Arg polymorphism & variant; p.Gly2019Ser mutation; p.A419V variant; p.Arg1628Pro variant; homozygous p.Ser1647Thr genotype (AA)	23115130; 17314670; 23624603; 17050822; 26234753; 18716801; 19854095; 18201193; 19672984; 20629711
FGF20	rs1989754, rs1721100, ss20399075; rs12720208; rs1721100 polymorphism & GG genotype; AGCCT haplotype, AGGGC haplotype	15122513; 26070653; 22342445; 24942208; 15122513
GBA	p.Leu444Pro, p.Asn386Lys, p.Pro428Ser, IVS2+1G>A, IVS9+3G>C, IVS10-9_10GT>AG and c.1309delG; p.Leu444Pro mutation; p.Asn188Ser, p.Pro201His, p.Arg257Gln and p.Ser271Gly	24997549; 20131388; 22387070
COMT	p.Met108Val polymorphism; p.Val158Met polymorphism	9503277; 23466833
DMT1	TT genotype and T allele of the 1254T>C polymorphism; CC haplotype	25817364; 21777657
HFE	p.Cys282Tyr SNP	12098643; 16824219
PARK16	rs947211, rs823128, rs823156, and rs11240572; rs947211 GG genotype	27174169; 21419001
PARK2	Heterozygous deletion of exon 2; p.Val56Glu, p.Cys212Tyr, deletion of exons 3/5 and 225delA	24729340; 12056932
PINK1	p.Gly411Ser heterozygotes; Heterozygous mutations	27807026; 16969854
SEMA5A	rs7702187 polymorphism; AC and AT haplotype	24706317; 18950607
SMPD1	c.996delC (p.Pro330fs), p.Leu302Pro, and p.Arg496Leu	26169695; 23535491
SNCA	rs356219 variant; dinucleotide polymorphism (REP1)	22349157; 18404644
TREM2	rs75932628 variant; p.Arg47His substitution	25936935; 23800361
ABCB1	g.116154T>C (rs28746504), g.117130A>G (rs2188524), g.117356C>G (rs34976462), g.117372T>C (rs3213619)	24572589
ACE	Deletion/insertion polymorphism	12084438
C9ORF72	Intermediate repeat copies	23845100
CARD15	p.Arg702Trp, p.Gly908Arg, p.Leu1007fs	17174426
pCD14	CD14 monocyte receptor gene polymorphism	16337421
CHCHD2	p.Pro2Leu variant	27626775
DGCR8	Hemizygous 22q11.2 deletions	24018986
DYRK1A	TT genotype derived from SNP rs8126696	27546826
GBA1	non-p.Asn370Ser GBA1 heterozygote mutation	22968580
GLIS1	Allele 'A' of rs797906 SNP	22759478
HLA-DRA	rs3129882 polymorphism	25319953
IL-10	Promoter−819 T>C polymorphism	22387064
IL-18	607C>A (rs1946518) polymorphism	21241672
MAO	MAO-A form loci haplotypes	7913737
MAPT	H1 haplotype	19912324
NOD2	p.Pro268Ser variant (rs2066842)	23651603
NOS2A	rs2255929 and rs1060826	16823855
HTRA2	IVS5+29T>A variant	21338583
P2X7	P2X7 receptor gene 1513A>C polymorphism	23648388
PARKIN	p.Ser167Asn heterozygotes	10511432
PLA2G6	p.Asp331Tyr mutation	22213678
Saitohin	rs62063857 polymorphism	25168738
SLC6A3	Genotype GG and allele G of the promoter SNP rs2652510	24211691
STK39	rs1955337 TT genotype	26469904
TFAM	rs2306604 GG genotype	19925850
TH	A rare heterozygous deletion of entire TH gene	20809526
UCHL1	AA genotype of UCHL1 p.Ser18Tyr SNP	22839974
VEGF	rs3025039 polymorphism	27481110
VMAT2	rs363324 at−-11.5 kb in the hVMAT2 promoter	27137201

Table 2 Epidemiological and environmental risk factors for PD

Epidemiological factors	PMID	Environmental factors	PMID
Age, male gender, and maternal age	19188574; 17884509; 21587290; 26952697	Duration of exposure to levodopa, along with age	23836370
Alpha-synuclein	15670652	Exposure to neuroleptics	10047927
Aromatase cytochrome P450 function	18063054	Hepatitis C virus infection	25608223; 26701382
Circadian rhythm dysfunction	27046648	Organophosphate or carbamate poisoning	27781262
Disturbance of the plasmatic rate of copper	23868039	Exposure to toxins other than pesticides /herbicides	21059511
Low plasma uric acid, selenium, iron, and zinc levels	18321759; 17443703; 24340079	Paraquat exposure	23150532
High serum lipid peroxidation rates	1641180	Farming	11437458
Rosacea	26999031	Exposure to well water	12140099
Cumulative impact of cardiovascular comorbidities	26576320	Exposure to trichloroethylene, perchloroethylene, and carbon tetrachloride	22083847
Constipation	27444575; 27234704	Hydrocarbon solvents	10980731
Anxiety	12784267	Multivitamin use	17408493
Depression	25995056; 24089392	Occupational lead exposure	17185278
Diabetes	19467671; 17251276	Nigral iron deposition	22288465
Hyperhomocysteinemia	24052451	Welding	11148228
Restless leg syndrome	27078707	Lead exposure	18607207
Non-apnea sleep disorders	28470725	Exposure to polychlorinated biphenyls	22906799
Ulcerative colitis	20584104	Exposure to pesticides	11437458
Head injury early in life, head trauma	25603768; 2057017; 2140099	Non-occupational pesticide and heavy metal exposure	19818671
Gaucher disease	20177787	Copper exposure	14706220
Metabolic syndrome	24955210	Manganese exposure	14706220

of PD. **Table 2** lists the previously identified PD-associated epidemiological and environmental factors. Gender and age, especially the maternal age, are known epidemiological factors that are significantly associated with PD. However, the association is conditional and could vary between individuals, given the contradictory reports of previous studies [20,21]. Many psychiatric disorders can also cause PD, such as anxiety and depression [22–24]. Since the different systems in our body interact and are linked with each other, many diseases in other systems, such as cardiovascular disorders and metabolic syndrome, can also be comorbidities or complications of PD, as listed in Table 2, and therefore risk factors for PD. Additionally, environmental exposure to pesticides, chemical solvents, drugs, and virus infections can affect gene expression and even

the ecological distribution of gut microbiota [25]. The complex interactions between these factors form a network that together regulates our biological systems and determines the course of PD.

More genetic and non-genetic risk factors are expected to be identified in future screening. However, their contribution to the pathogenesis of PD is too complex to investigate without big data collected from PD patients and healthy people as controls. Two complementary approaches are now available for the screening of PD risk factors or biomarkers. One approach is the cross-sectional cohort study designed to collect medical data from a group of people and then to identify statistically significant features of disease risk factors [26,27], while the other approach is the longitudinal personalized study

of individuals to identify patterns associated with disease and health status among these individuals [28]. The former approach can identify important features common to a population, but a one-size-fits-all threshold may not be accurate for individuals when applying statistically averaged indicators to disease diagnosis, like applying the same blood pressure thresholds to the diagnosis of high hypertension. The latter approach assesses the individual's health status based on personalized reference data, which could be more accurate than the former approach for a personalized diagnosis, although more longitudinal data on individuals need to be collected for the latter approach.

Lifestyle changes for the prevention of Parkinson's disease

Compared to genetic and environmental factors, lifestyle can be adjusted more easily for the prevention of disease and improvement of health. As shown in Table 3, two types of lifestyle behaviours have been found to positively and negatively affect PD. Although smoking is a risk factor for cancer, especially lung cancer, it could be a preventive factor for PD [29,30]. In addition, the consumption of coffee, tea, wine, *etc.* could be helpful for the prevention of PD as well. Although the relationships between lifestyles and diseases are complicated by their interaction with genetic and environmental factors, negative lifestyle behaviours should be adjusted to reduce the risk of PD. Particularly, in the era of elderly society, actively changing lifestyles for the prevention of disease is a better strategy for healthcare than traditional clinical intervention, which is cost prohibitive and requires more labour and medical resources. Lifestyle management for high-risk populations is an efficient way to prevent PD [31].

To unravel relationships between lifestyle and disease prevention, exclusive use of biomedical data is inadequate, and mining data from social networks will be important. Differences between "translational bioinformatics", "translational biomedical informatics", and "translational informatics" are related to the data types analyzed. Bioinformatics generally focuses on data at the molecular level, such as genome, transcriptome, proteome, and metabolome data, whereas biomedical informatics also involves cell/tissue imaging data, patient data, and the public health data. As noted, translational informatics will cover a wider scope of data relative to the other two methodologies, as it includes not only biomedical data but also social network data associated with lifestyle information.

Driver player and biomarker discovery for personalized medicine

Biomarkers are a class of indicators that are able to predict changes in biological systems and provide specific signatures for disease diagnosis, prognosis, or treatment [7]. In recent decades, many PD-related biomarkers, including molecules, images, clinical symptoms, and physiology, have been identified for monitoring the occurrence and progression of this complex disease.

As illustrated in Table 4, biological molecules, such as genes, RNAs, proteins, and metabolites, play important roles in PD evolution. For example, cerebro-spinal fluid (CSF) α-synuclein was one of the well-studied proteins implicated in PD pathogenesis, and its genetic variability was a prognostic marker for PD, PD with dementia, and dementia with Lewy bodies [32]. Ritz et al. [33] demonstrated that α-synuclein genetic variants were associated with the development of faster motor symptoms in idiopathic PD. In addition, Ballard et al. [34] reported that CSF α-synuclein had the potential for diagnosing PD and related dementias. Mollenhauer et al. [26] found that CSF α-synuclein was also a useful indicator in PD patients undergoing dopamine replacement therapy. Moreover, plasma and skin nerve α-synuclein is valuable in predicting PD cognitive impairment and idiopathic PD, respectively [35,36]. Another key player, CSF β-amyloid 1–42, was a powerful predictor of the progression of cognitive impairment, dementia, and dopa-resistant gait in PD. For example, a lower level of CSF β-amyloid 1–42 was common in advanced PD patients with cognitive decline and could be used to predict cognitive impairment in newly diagnosed PD [37]. Alves et al. [38] indicated that the CSF levels of β-amyloid 1–42 were lower in PD patients with dementia. The abnormal expression of this protein increased the risk of dementia development, which was used for the early prognosis of PD dementia [38]. Moreover, a decrease in β-amyloid 1–42 was also involved in the pathology of dopa-resistant gait in early PD [39].

In contrast to molecular biomarkers, imaging and clinical symptoms are often directly used clinically for PD investigation. With the development of medical imaging techniques, PET imaging, quantitative EEG, and single photon emission computed tomography (SPE-CT) have been widely used to screen key signatures to predict the progression of dementia, the severity of fatigue, and dopaminergic responsiveness in PD patients [40–43]. Clinical symptoms, on the other hand, call attention to an early diagnosis of PD and its associated phenotypes. For example, episodic anxiety was found to be more specific for the anxiety subtypes in PD than the persistent anxiety. Episodic anxiety was a significant factor related to PD severity and duration [44]. Based on a case-control study, Pradhan et al. [45] uncovered that characteristics in grip force modulation, *e.g.*, force and movement quality, were sensitive measurements in detecting early PD and tracking the clinical progression of PD patients. A circadian change in core body

Table 3 Positive and negative lifestyles for PD

Positive lifestyle	PMID	Negative lifestyle	PMID
Coffee, tea	29057010; 18607207; 18398911; 23377703; 18156141; 14579122; 11456310	Habitual longer sleep duration	16495472
Smoking	29057010; 18607207; 23377703; 18156141; 14579122; 10230846; 10970060; 14706220; 12956856	Excessive daytime sleepiness	16275833
Fresh vegetables, fruit, and herbs	29081890	High carbohydrates	14579122
Nuts and seeds, non-fried fish	29081890	High intake of animal fat	10970060
Spices, olive oil, and coconut oil	29081890	Dairy products	18607207; 23377703
Wine	29081890	Working night shifts	16495472

Table 4 Literature-reported biomarkers for diagnosis, prognosis, and treatment of PD

Group	Item	Classification and function	PMID
Molecule	*SNCA* genetic variability	Prognosis for PD, PD with dementia, and dementia with Lewy bodies, and prediction for the progression of faster motor symptom in idiopathic PD	27091628; 22615757
	α-synuclein	Diagnosis and prognosis for PD cognitive impairment, dementias, and dopamine replacement therapy	20962288; 29030452; 28643805; 25332447
	β-amyloid 1-42	Prediction for cognitive impairment, dementia, and the progression of dopa-resistant gait in PD	26330275; 24748671; 28330963; 20720189
	Urate	Prediction of the clinical decline in PD progression, and a promising target for PD disease-modifying therapy	19822770; 18413464
	COMT Val158Met and *BDNF* Val66Met variants	Prediction for cognitive impairment in PD	28520803
	Parkin human gene mutations	Prediction for early-onset PD	20558392
	VEGF, *PlGF*, *sVEGFR-2*, and *Ang2* expression	Prediction for the angiogenesis in PD	26511451
	HLF, *E2F1*, and *STAT4* expression	Diagnosis for PD	23284986
	IGF-1 expression	Diagnosis and risk prediction for PD	21747033
	EGF expression	Prediction for cognitive impairment in PD	21520231
	PARK2 and *PARK8* expression	Diagnosis and risk prediction for PD	26760142
	miR-141, miR-214, miR-146b-5p, and miR-193a-3p	Diagnosis for early onset of PD	26631297
	miR-103a-3p, miR-30b-5p, and miR-29a-3p	Prognosis for L-dopa-treated PD patients	25596505
	CSF tau protein, total-tau/Aβ-42	Prediction for neurodegeneration severity in PD	22065209
	Lipopolysaccharide binding protein	Prediction for gastrointestinal inflammation in PD	26388718
	Apolipoprotein A1	Prediction for the risk of PD	23447138
	DJ-1 isoforms	Prediction for the late-stage of PD	23233873
	CSF Flt3 ligand, fractalkine/Aβ(1-42)	Diagnosis for PD and prognosis for PD severity and progression	21400565
	Dopamine deficiency	Prediction for cognitive impairment of PD	28520803
	Putaminal dopamine turnover	Prediction for later motor complications in PD	26718573
	CSF peptides	Diagnosis for PD and prediction for the progression of PD	25556233

Table 4 (continued)

Group	Item	Classification and function	PMID
Image	Fluorodeoxyglucose-PET statistical parametric mapping	Prediction for the progression of dementia in PD	29453242
	Dopaminergic and cholinergic PET	Prediction for the severity of fatigue in PD	26683744
	Quantitative EEG	Prediction for dementia incidence in PD	21633128
	Striatal asymmetry index by SPECT	Prediction for dopaminergic responsiveness in PD patients	21406999
Clinical symptom and physiology	Episodic anxiety	Diagnosis for PD anxiety subtypes	29528763
	Grip force modulation	Diagnosis for the early PD	25476717
	Core body temperature	Prediction for the depression in PD	18332645
	Dementia	Prediction for the survival of PD patients	22213411
	Brain volume or thickness	Prediction for cognitive impairment of PD	28520803
	Resting tremor	Prediction for the probability of levodopa-induced dyskinesia in PD	21825240
	REM sleep behavior disorder	Prediction for cognitive impairment in PD	17984452
	Body sway velocity and ability	Prediction for the falls in PD	25690532
	Electrophysiologic subthalamic nucleus length	Prediction for the outcome of deep brain stimulation in PD	28982098
	Connectivity between stimulation site and subthalamic nucleus	Prediction for the outcome of deep brain stimulation in PD	28586141
	Retinal changes	Diagnosis for early PD	27586255
	Orthostatic hypotension	Prediction for motor decline in early PD	27639815
	Sympathetic skin response and heart rate variability	Diagnosis for autonomic disorders in PD patients	28471954
	Non-motor symptom burden and somatic affective symptoms	Prediction for the fatigue in PD	26683744
Others	Caffeine and its metabolites	Diagnosis for early PD	29298852
	Sleep	Prediction for the progression of early PD	27164658
	25-hydroxy-vitamin D3	Prediction for the risk of PD	24068787
	Midbrain iron content	Prediction for PD status	18172063
	8-hydroxydeoxyguanosine	Prediction for PD stages and progression	15781836

Note: Molecule types includes gene, RNA, protein, and metabolite; CSF, cerebrospinal fluid; PET, positron emission tomography; EEG, electroencephalograph; SPECT, single photon emission computed tomography; REM, rapid eye movement.

temperature, *i.e.*, rectal temperature, was detected in PD patients with depression, suggesting its possibility in predicting PD depression [46]. Typical clinical symptoms could be evaluated for PD prognosis tracking in addition to diagnosis. Willis et al. [47] showed that dementia was a prevalent trait in PD patients, which strongly affected the survival of PD patients and could increase the chance of mortality. In addition to clinical symptoms, some physiological features were connected with the functional alterations observed in PD. For example, the length of the electrophysiologic subthalamic nucleus and

the connectivity between the stimulation site and subthalamic nucleus could predict the outcome of deep brain stimulation in PD [48,49]. Brain volume or thickness could also be used as a parameter to recognize cognitive impairment during PD development [50].

With the coming age of big data and digitalized medicine, more novel and important functional components in our biological systems will be discovered, such as how gut flora dysbiosis can affect brain function and how it is associated with PD through the microbiota-gut-brain axis. Therefore gut flora dysbiosis can be used as a new type of biomarker for PD [51].

Data integration and modelling for translational informatics of PD

PD biomedical data standardization and integration

Big PD biomedical data are diverse and could include the following data types. (1) The first one is different omics data, such as genomic data on genetic structure, variant susceptibility to PD, and the transcriptomic, as well as proteomic and metabolomic data characterization of the abnormal states of PD. (2) The second is neuroimaging data, such as MRI for brain functional structure, as well as PET-CT and SPE-CT for altered brain structures in PD patients. (3) The third type of data are physiological signals, including EEG and electrocardiography (ECG) to reflect PD patients' cognitive impairment or other clinical features [52,53]. (4) The fourth type of data are information from EHR or electronic medical records (EMR). These include patient's demographic data, results of clinical laboratory tests, medical history, use of specific medications, and other clinical phenotype data. (5) Finally, the last type of data are epidemiological data on lifestyle, environment, or social network information.

These big PD biomedical data could be static or dynamic and can reflect the development of PD from early prodromal symptoms to the latter clinical stages. As presented in Figure 3, big PD data have several characteristics that differ from those of other types of big data such as business, market, and social network data. Notably, data privacy is important for patients ethically, and the data need to be transformed before they can be accessed by users and researchers. The heterogeneity of PD is caused by the interaction between multiple pathogenic factors, such as genetics, lifestyle, and environment. In addition, these data could be collected from different platforms and stages of different patients. The PD data listed above could be collected at different levels ranging from molecular to cellular, tissue, or individual levels at different time points. These properties make standardization and integration very challenging. The challenges for PD data integration could include the following three aspects.

Challenge 1: *diversity, standardization, and sharing of big PD biomedical data*

The genesis and progression of PD is caused by the complex interactions between genetics, environmental factors and lifestyles and the PD phenotypes are therefore very diverse and heterogeneous [54,55]. To understand mechanism of PD at the systems biological level, the different omics data need to be integrated and standardized for sharing and modelling. For the sharing of big PD biomedical data, two issues need to be considered. The first issue is data privacy preservation [56-59]. Although many algorithms have been developed for the protection of patient's genome information, further efforts are needed to preserve the patients' information at the phenotype and family levels [60-63]. The second issue is the development of ontology for the standardization of PD data, which could classify and standardize the PD specific concepts and synonyms and promote the sharing and integration of big biomedical data on PD.

Challenge 2: *databases for big PD biomedical data and knowledge*

A database and a knowledge base for diverse PD data are needed for the modelling and understanding of the pathogenesis and progression of PD. Table 5 lists the existing PD databases. PDGene is a comprehensive online resource of potential risk loci in PD [64]. After data from all the published articles and genome-wide association studies (GWAS) were extracted, deep meta-analyses were performed on millions of polymorphisms from different GWAS datasets or PD-related studies. A total of 11 loci, *e.g.*, *GBA*, *LRRK2*, *MAPT*, *PARK16*, and
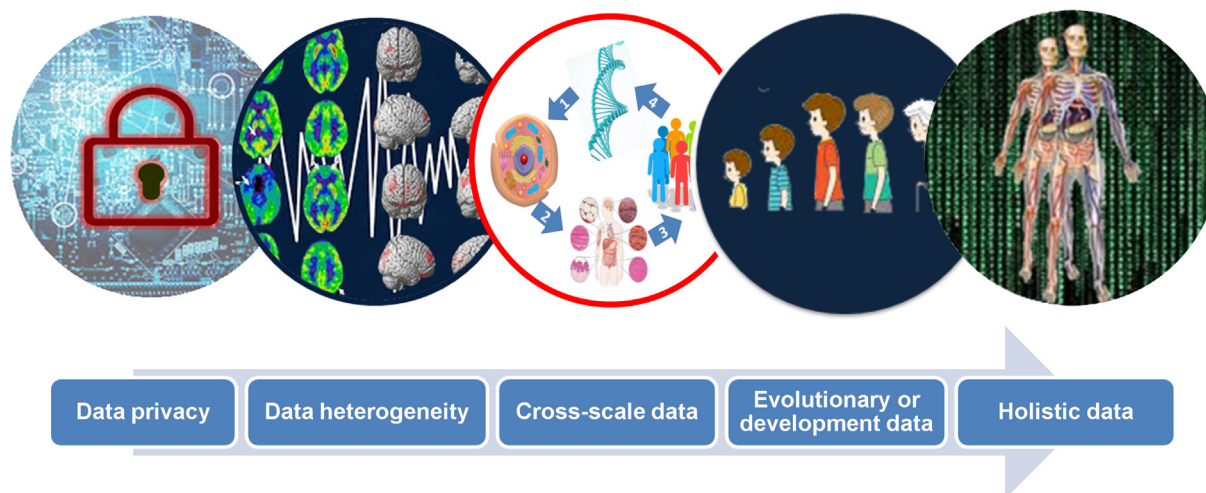

Figure 3   Diverse data types and big data challenges

Table 5   The currently-available PD databases

Name	Description	URL	PMID
PDGene	A comprehensive database for collection and meta-analysis of PD-related risk loci from published papers	http://www.pdgene.org/	22438815
ParkDB	A queryable database of gene expression in PD	http://www2.cancer.ucl.ac.uk/Parkinson_Db2/	21593080
PDmutDB	A comprehensive database of genetic mutations associated with PD development	http://www.molgen.ua.ac.be/PDmutDB/	22581678
PDbase	A comprehensive database of PD-related genes, genetic variations, and other functional components for PD causation understanding	http://bioportal.kobic.re.kr/PDbase/	19958497
PPMI	An international and multi-center study, Parkinson's Progression Markers Initiative	http://www.ppmi-info.org/	21930184

SNCA, were significant genome-wide for PD risk evaluation. ParkDB is another database aimed at recording key molecular events during PD development [65]. It contains a large number of re-analysed and annotated microarray datasets, which are advantageous for screening expression signatures associated with PD under different biological backgrounds. PDmutDB is a PD mutation database that comprises information on all known mutations in the genes associated with PD development [66]. Through expressed sequenced tags (ESTs) on substantia nigra tissues from healthy and PD populations, PDbase was built to capture PD-related genes and genetic variations [67]. In addition, the database integrated several valuable resources for PD annotation and provides information such as mitochondrion proteins, microRNA-gene regulations, structural variations in PD-related genes, and pathways/networks within protein–protein interactions to better understand the causes of PD. An international and multi-centre study, PPMI, also collected diverse data from PD patients for future biomarker discovery and personalized PD therapy [68]. Although many databases have been established, with more digital data from PD patients and related resources available, we could collect big PD biomedical data, especially PD-associated phenotype data, to conduct studies to obtain a holistic description and mechanism of PD.

Challenge 3: *cross-level and dynamic integration of PD biomedical data*

Many levels exist between genotype and PD clinical symptom phenotype, such as the molecular phenotype and cellular phenotype, and physiological signals could also be a type of phenotype. Therefore, the relationship between genotype and a patient's clinical phenotype is very complex. At present, most
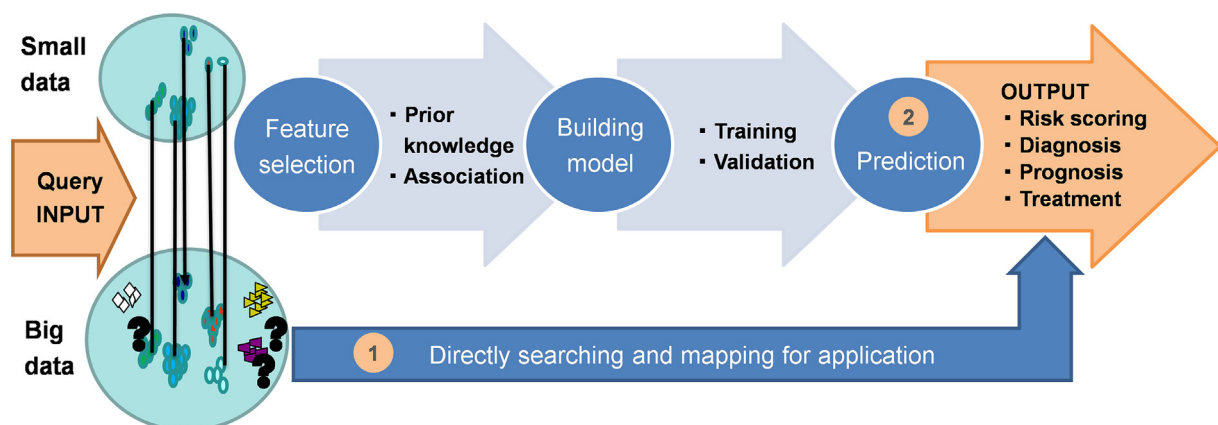
Figure 4   Big data model for precision prediction

of the PD data and information at all these levels are isolated from each other and need to be interlinked and integrated. In the time dimension, these data can be ordered based on pathogenesis and progression. Traditionally, data at these different levels are often statistically averaged and reasoned for correlation studies; however, these methods often average the patterns in subgroups of the studied samples. Paired data for all the levels between the genotype and disease phenotype will be essential to the precision modelling of the disease systems, and if the paired data are collected in a time series, then the PD progression and trajectory could be modelled. The Cancer Genome Atlas (TCGA) for cancer research is a typical paradigm that could be applied to PD data integration in the future to obtain cross-level and dynamic integration of PD data.

Big PD data mining and modelling for translational application

As shown in **Figure 4**, when small data are used for PD modelling, some complex patterns cannot be represented in this small data space; thus, when a model trained from a small data set is applied to a big data space, the model will unlikely be successful. With big PD biomedical data available, we will have the chance to use these data to model and mine the knowledge and patterns hidden in these big data, and some questions that could not be answered before could now be investigated. The following three modelling challenges are expected when translating big PD biomedical data to clinical application.

Challenge 4: holistic and systems-level modelling and mechanism-based key player identification
Without sufficient data for modelling PD pathogenesis and progression, we can only partly address the complex PD "elephant". Holistic and systems-level characterization of the PD mechanism is necessary to understand its complexity and heterogeneity. The systems-level identification of key players, such as biomarkers for classifying PD and risk factors for high-risk population screening, will be a challenge for future PD translational informatics. Compared to traditional disease-gene recognition, systems-level discovery of genes, pathways, modules or sub-networks that drive systems to change from a healthy to a disease state will be the objectives of big data-based modelling.

Table 6 lists previously reported PD-associated pathways, as based on our experience in cancer research, it is often easier to find common pathways for complex diseases than to find common disease genes [69,70].

Challenge 5: modelling of PD dynamic progression and systems-level control of PD progression
Since complex PD is the product of a dynamic interaction between the patient's genetics, environment and lifestyle, the cause and course of PD are dynamically changed. With dynamic information from the human body, such as routine blood testing and the real-time collection of physiological signals [12], modelling of the dynamic evolution of PD is possible, and the identification of the key hubs and connections in these dynamic systems will be a challenge but opportunity for rational drug design or lifestyle changes to control the development of PD [71,72].

Table 6   Biological pathways associated with PD pathogenesis and molecular mechanisms

Pathway	Function	PMID
Ubiquitin-proteasomal pathway	Mutations of *mUchL1* and parkin proteins in this pathway could cause familial PD	11881748
Glia pathway	This pathway protects neurons and provides novel therapeutic strategies for PD treatment	16399902
pRb/E2F cell-cycle pathway	This pathway could medicate cell death in PD	17360686
JNK pathway	This pathway leads to the pathogenesis of familial PD by linking the malfunction of mutations in *DJ-1*, *PARKIN*, *PINK1*, and alpha-SYNUCLEIN to apoptosis	18399358
MAPK signaling	Mutations in *LRRK2* causes G2019S substitution, which could affect MAPK signaling cascades and contribute to PD	17385669
Calpain/cdk5/p25 pathway	p25/p35 immunoreactivity ratio is increased and cdk5/p25 pathway is involved in the process of neuronal loss in PD patients	17977053
PI3K/Akt pathway	Caffeine prevents SH-SY5Y cells from cell death by activating this pathway	18201823
Akt/Nrf2/glutathione pathway	DBL prevents the toxicity of PD-related neurotoxin 6-hydroxydopamine via this pathway	23959789
Estrogen receptor *beta-PI3K*/Akt pathway	This pathway is associated with the cytoprotective effects of tocotrienol in PD	24768803
Akt/Nrf2 and Akt/*CREB* pathway	In *in vitro* and *in vivo* PD models, 11-dehydrosinulariolide prevents 6−OHDA−induced damage on SH−SY5Y cell via this pathway	27763504
Autophagy-lysosome pathway	Alpha-synuclein aggregation or mutation in *ATP13A2* results in dysfunction of this pathway, and impairment of this pathway is associated with PD development	18187492
Wnt signaling pathway	Key regulators in this pathway, *e.g.*, DVL1-3, could interact with LRRK2 and play pathogenic roles in PD	19625296
17β-estradiol signaling	This pathway interacts with multiple signaling molecules and mediates the neuroprotective effects of sex steroids in PD	22387674
Dopaminergic neurotransmission pathway	Inhibiting *COMT* in this pathway could prevent the metabolism of dopamine, which is helpful for PD drug development	22622642
Axonal guidance signaling	Polymorphism of genes in this pathway contributed to PD pathogenesis	17571925
RAGE/TNF-alpha pathway	*S100B* is over-expressed in post-mortem substantia nigra of PD patients, and ablation of *S100B* prevents *MPTP*-induced toxicity via this pathway	23169921
PINK1/parkin pathway	*AF-6* is involved in PD pathogenesis by modulating this pathway	23393160
Daf-2/Daf-16 insulin like signaling pathway	Ida-1 in this pathway is a common modulator between PD and diabetes	25469508
p53 pathway	The downregulation of *PCNA* mediates 1-methyl-4-phenylpyridinium-induced oxidative damage in PD via this pathway	26677001
Phosphatidylcholine biosynthesis I and *PPAR* signaling, *mTOR* and *CD28* signaling pathways	These paired pathways could accurately classify PD and healthy samples, and hold the potential as biomarkers for PD diagnosis and treatment	27146810

Note: DBL, 3,4-dihydroxybenzalacetone; MPTP, 1-methyl-4-phenyl-1,2,4,6,-tetrahydropyridine; 6-OHDA, 6-hydroxydopamine; MPP, 1-methyl-4-phenylpyridinium.

Challenge 6: general rule discovery for basic research and prevention of PD

Big data make artificial intelligence, including deep learning and reinforcement learning, applicable to the analysis of big data and PD studies [73,74]. Furthermore, knowledge of PD is accumulating and could be used to improve predictions [75]. However, the discovery of general rules for the molecular mechanism of disease is still very necessary to study complex systems. Our previous study discovered the rich-get-richer rule for a new gene's functional evolution [76], and for disease progression and prevention, the discovery of general rules from big biomedical data will be a complementary objective to personalized and precision PD medicine.

We face two additional challenges in the translational application of findings from big data to PD clinical management and healthcare.

Challenge 7: screening of populations at high risk of PD

Integrating genetic susceptibility and environmental and lifestyle factors together to build a systems model for the precision screening of populations at high risk for PD will be essential for the early prevention and intervention of PD.

Challenge 8: PD sub-population searching for personalized treatments

PD patients can be treated with levodopa or other dopamine replacement drugs, and surgical approaches, including pallidotomy and thalamotomy, could be alternatives. The responses and side effects of these therapies can be personalized to each patient [77,78]. Big data could also provide a direct search and mapping method for the clinical decision of which treatment to use, as shown in **Figure 4**. However, identifying suitable sub-populations for efficient treatments is always a challenge for clinical application.

Perspectives on future translational PD informatics

The driving forces for the translational informatics study of PD come from technological, scientific and social aspects. Technically, genome variants, gene expression and epigenetic
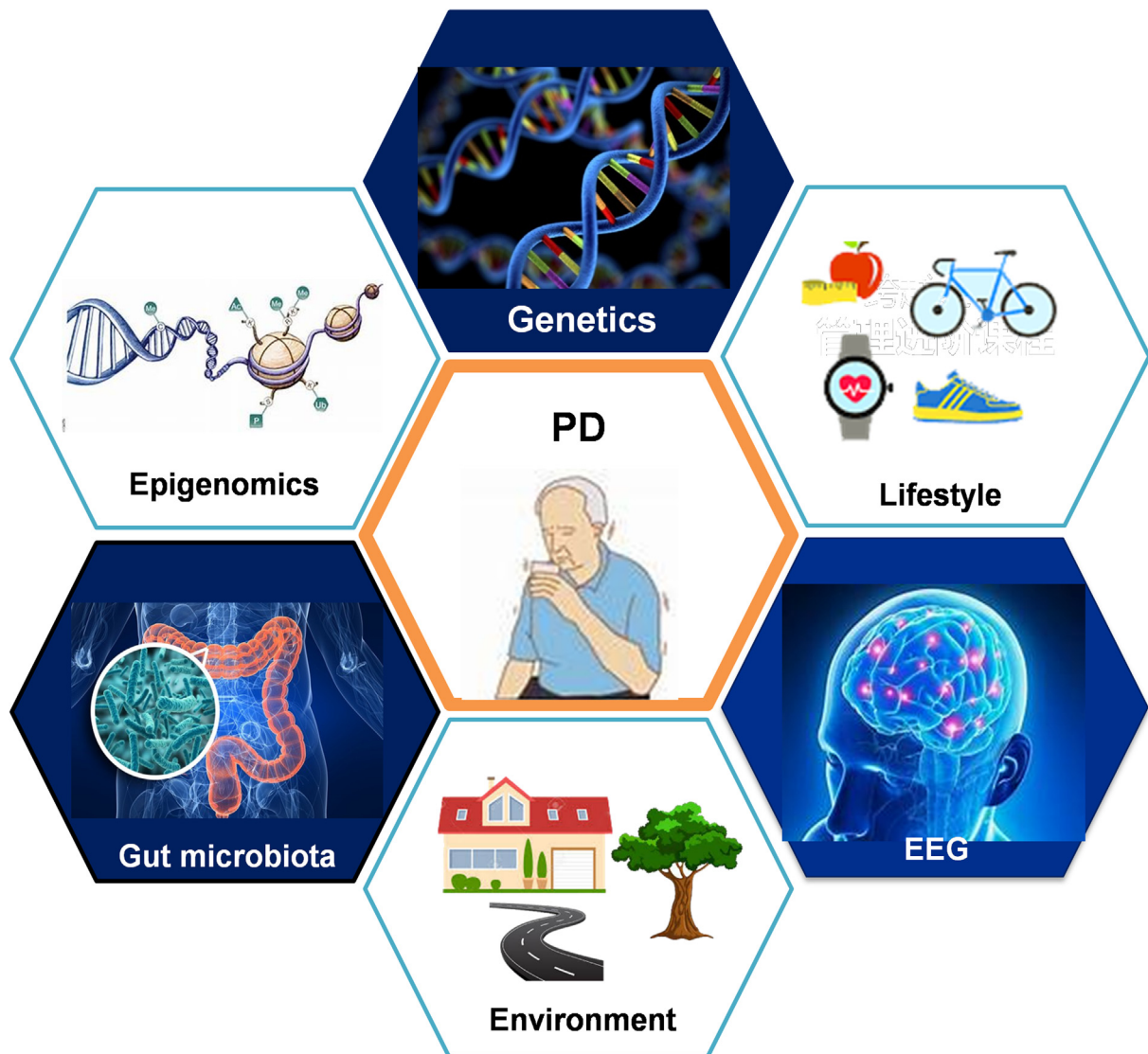

Figure 5   From personalized data to systems healthcare of PD patients

alterations, *etc.* could be measured by advanced next-generation sequencing technologies. Clinical laboratory tests may be easily performed by point-of-care tests in a direct-to-consumer mode. Physiological signals could be detected in real time by the combination of wearable sensors, smart phones and cloud computing, and everyone, including healthy people, family members, nurses, medical doctors and data analysts, could be linked via the internet in the cloud to manage the data in a crowdsourcing model.

Scientifically, the interactions between genetics, lifestyle and physiological signals as well as the microbiota and the environment are deepening our knowledge and understanding of PD (**Figure 5**). Systems biology and evolutionary medicine-level modelling of these interactions are becoming the paradigm to investigate complex diseases such as cancer and NDDs, including PD. Recent genetic editing methods, findings regarding brain-gut connections and studies on the diverse molecular mechanisms of PD pathogenesis all accelerate basic PD research discoveries for clinical applications.

Regarding the social and economic aspects, the ageing society and the considerable cost of the clinical management of PD urgently demand improved prevention and prediction of PD, and all governments are promoting the market of healthcare, especially for senile diseases such as AD and PD. By addressing the three challenges to PD data integration described above, translational informatics for PD studies will have considerable opportunities for scientific discovery and healthcare applications.

Competing interests

The authors have declared no competing interests.

Acknowledgments

References

[1] Hoyert DL, Xu J. Deaths: preliminary data for 2011. Natl Vital Stat Rep 2012;61:1–51.

[2] Hawkes CH, Del Tredici K, Braak H. A timeline for Parkinson's disease. Parkinsonism Relat Disord 2010;16:79–84.

[3] Abeliovich A, Rhinn H. Parkinson's disease: guilt by genetic association. Nature 2016;533:40–1.

[4] Sampson TR, Debelius JW, Thron T, Janssen S, Shastri GG, Ilhan ZE, et al. Gut microbiota regulate motor deficits and neuroinflammation in a model of Parkinson's disease. Cell 2016;167:1469–80.e12.

[5] Bellou V, Belbasis L, Tzoulaki I, Evangelou E, Ioannidis JP. Environmental risk factors and Parkinson's disease: an umbrella review of meta-analyses. Parkinsonism Relat Disord 2016;23:1–9.

[6] Santangelo G, Piscopo F, Barone P, Vitale C. Personality in Parkinson's disease: clinical, behavioural and cognitive correlates. J Neurol Sci 2017;374:17–25.

[7] Lin Y, Qian F, Shen L, Chen F, Chen J, Shen B. Computer-aided biomarker discovery for precision medicine: data resources, models and applications. Brief Bioinform 2019;20:952–75.

[8] Lin Y, Chen J, Shen B. Interactions between genetics, lifestyle, and environmental factors for healthcare. Adv Exp Med Biol 2017;1005:167–91.

[9] Chen J, Lin Y, Shen B. Informatics for precision medicine and healthcare. Adv Exp Med Biol 2017;1005:1–20.

[10] Chen J, Qian F, Yan W, Shen B. Translational biomedical informatics in the cloud: present and future. Biomed Res Int 2013;2013:658925.

[11] Shen B, Teschendorff AE, Zhi D, Xia J. Biomedical data integration, modeling, and simulation in the era of big data and translational medicine. Biomed Res Int 2014;2014:731546.

[12] Bai J, Shen L, Sun H, Shen B. Physiological informatics: collection and analyses of data from wearable sensors and smartphone for healthcare. Adv Exp Med Biol 2017;1028:17–37.

[13] Goedert M. Familial Parkinson's disease. The awakening of alpha-synuclein. Nature 1997;388:232–3.

[14] Yang Y, Xu C, Liu X, Xu C, Zhang Y, Shen L, et al. NDDVD: an integrated and manually curated neurodegenerative diseases variation database. Database (Oxford) 2018;2018:bay018.

[15] Rideout HJ. Neuronal death signaling pathways triggered by mutant LRRK2. Biochem Soc Trans 2017;45:123–9.

[16] Lee BD, Dawson VL, Dawson TM. Leucine-rich repeat kinase 2 (LRRK2) as a potential therapeutic target in Parkinson's disease. Trends Pharmacol Sci 2012;33:365–73.

[17] Tan EK, Schapira AH. LRRK2 as a therapeutic target in Parkinson's disease. Eur J Neurol 2011;18:545–6.

[18] Chen J, Sun M, Shen B. Deciphering oncogenic drivers: from single genes to integrated pathways. Brief Bioinform 2015;16:413–28.

[19] Billingsley KJ, Bandres-Ciga S, Saez-Atienzar S, Singleton AB. Genetic risk factors in Parkinson's disease. Cell Tissue Res 2018;373:9–20.

[20] Gardener H, Gao X, Chen H, Schwarzschild MA, Spiegelman D, Ascherio A. Prenatal and early life factors and risk of Parkinson's disease. Mov Disord 2010;25:1560–7.

[21] Currie LJ, Harrison MB, Trugman JM, Bennett JP, Swerdlow RH, Manning CA, et al. Maternal age is not a risk factor for Parkinson's disease. J Neurol Neurosurg Psychiatry 2001;71:130–1.

[22] Witt I, Ganjavi H, MacDonald P. Relationship between freezing of gait and anxiety in Parkinson's disease patients: a systemic literature review. Parkinsons Dis 2019;2019:6836082.

[23] van der Velden RMJ, Broen MPG, Kuijf ML, Leentjens AFG. Frequency of mood and anxiety fluctuations in Parkinson's disease patients with motor fluctuations: a systematic review. Mov Disord 2018;33:1521–7.

[24] Fontoura JL, Baptista C, Pedroso FB, Pochapski JA, Miyoshi E, Ferro MM. Depression in Parkinson's disease: the contribution from animal studies. Parkinsons Dis 2017;2017:9124160.

[25] Mayer EA, Tillisch K, Gupta A. Gut/brain axis and the microbiota. J Clin Invest 2015;125:926–38.

[26] Mollenhauer B, Caspell-Garcia CJ, Coffey CS, Taylor P, Shaw LM, Trojanowski JQ, et al. Longitudinal CSF biomarkers in patients with early Parkinson disease and healthy controls. Neurology 2017;89:1959–69.

[27] Simuni T, Siderowf A, Lasch S, Coffey CS, Caspell-Garcia C, Jennings D, et al. Longitudinal change of clinical and biological measures in early Parkinson's disease: Parkinson's progression markers initiative cohort. Mov Disord 2018;33:771–82.

[28] Price ND, Magis AT, Earls JC, Glusman G, Levy R, Lausted C, et al. A wellness study of 108 individuals using personal, dense, dynamic data clouds. Nat Biotechnol 2017;35:747–56.

[29] Gigante AF, Defazio G, Asabella AN, Superbo M, Ferrari C, Liuzzi D, et al. Smoking in patients with Parkinson's disease:

preliminary striatal DaT-SPECT findings. Acta Neurol Scand 2016;134:265–70.

[30] Rahal Z, El Nemr S, Sinjab A, Chami H, Tfayli A, Kadara H. Smoking and lung cancer: a geo-regional perspective. Front Oncol 2017;7:194.

[31] Lim SY, Tan AH, Fox SH, Evans AH, Low SC. Integrating patient concerns into Parkinson's disease management. Curr Neurol Neurosci Rep 2017;17:3.

[32] Guella I, Evans DM, Szu-Tu C, Nosova E, Bortnick SF, SNCA Cognition Study Group, et al. Alpha-synuclein genetic variability: a biomarker for dementia in Parkinson disease. Ann Neurol 2016;79:991–9.

[33] Ritz B, Rhodes SL, Bordelon Y, Bronstein J. alpha-Synuclein genetic variants predict faster motor symptom progression in idiopathic Parkinson disease. PLoS One 2012;7:e36199.

[34] Ballard CG, Jones EL. CSF alpha-synuclein as a diagnostic biomarker for Parkinson disease and related dementias. Neurology 2010;75:1760–1.

[35] Wood H. Parkinson disease: plasma alpha-synuclein – a potential marker of cognitive impairment in Parkinson disease. Nat Rev Neurol 2017;13:450.

[36] Sharma SK. Skin nerve alpha-synuclein deposits: a biomarker for idiopathic Parkinson disease. Neurology 2014;83:1582.

[37] Terrelonge M, Marder KS, Weintraub D, Alcalay RN. CSF beta-amyloid 1–42 predicts progression to cognitive impairment in newly diagnosed Parkinson disease. J Mol Neurosci 2016;58:88–92.

[38] Alves G, Lange J, Blennow K, Zetterberg H, Andreasson U, Forland MG, et al. CSF Abeta42 predicts early-onset dementia in Parkinson disease. Neurology 2014;82:1784–90.

[39] Rochester L, Galna B, Lord S, Yarnall AJ, Morris R, Duncan G, et al. Decrease in Abeta42 predicts dopa-resistant gait progression in early Parkinson disease. Neurology 2017;88:1501–11.

[40] Pilotto A, Premi E, Caminiti SP, Presotto L, Turrone R, Alberici A, et al. Single-subject SPM FDG-PET patterns predict risk of dementia progression in Parkinson disease. Neurology 2018;90:e1029–37.

[41] Chou KL, Kotagal V, Bohnen NI. Neuroimaging and clinical predictors of fatigue in Parkinson disease. Parkinsonism Relat Disord 2016;23:45–9.

[42] Klassen BT, Hentz JG, Shill HA, Driver-Dunckley E, Evidente VG, Sabbagh MN, et al. Quantitative EEG as a predictive biomarker for Parkinson disease dementia. Neurology 2011;77:118–24.

[43] Contrafatto D, Mostile G, Nicoletti A, Raciti L, Luca A, Dibilio V, et al. Single photon emission computed tomography striatal asymmetry index may predict dopaminergic responsiveness in Parkinson disease. Clin Neuropharmacol 2011;34:71–3.

[44] Broen MPG, Leentjens AFG, Hinkle JT, Moonen AJH, Kuijf ML, Fischer NM, et al. Clinical markers of anxiety subtypes in Parkinson disease. J Geriatr Psychiatry Neurol 2018;31:55–62.

[45] Pradhan S, Scherer R, Matsuoka Y, Kelly VE. Grip force modulation characteristics as a marker for clinical disease progression in individuals with Parkinson disease: case-control study. Phys Ther 2015;95:369–79.

[46] Suzuki K, Miyamoto T, Miyamoto M, Kaji Y, Takekawa H, Hirata K. Circadian variation of core body temperature in Parkinson disease patients with depression: a potential biological marker for depression in Parkinson disease. Neuropsychobiology 2007;56:172–9.

[47] Willis AW, Schootman M, Kung N, Evanoff BA, Perlmutter JS, Racette BA. Predictors of survival in patients with Parkinson disease. Arch Neurol 2012;69:601–7.

[48] Boex C, Tyrand R, Horvath J, Fleury V, Sadri S, Corniola M, et al. What is the best electrophysiologic marker of the outcome of subthalamic nucleus stimulation in Parkinson disease? World Neurosurg 2018;120:e1217–24.

[49] Shenai MB, Patel DM, Romeo A, Whisenhunt JD, Walker HC, Guthrie S, et al. The relationship of electrophysiologic subthalamic nucleus length as a predictor of outcomes in deep brain stimulation for Parkinson disease. Stereotact Funct Neurosurg 2017;95:341–7.

[50] Caspell-Garcia C, Simuni T, Tosun-Turgut D, Wu IW, Zhang Y, Nalls M, et al. Multiple modality biomarker prediction of cognitive impairment in prospectively followed de novo Parkinson disease. PLoS One 2017;12:e0175674.

[51] Smith LM, Parr-Brownlie LC. A neuroscience perspective of the gut theory of Parkinson's disease. Eur J Neurosci 2019;49:817–23.

[52] Mochizuki H, Ishii N, Shiomi K, Nakazato M. Clinical features and electrocardiogram parameters in Parkinson's disease. Neurol Int 2017;9:7356.

[53] Guntekin B, Hanoglu L, Guner D, Yilmaz NH, Cadirci F, Mantar N, et al. Cognitive impairment in Parkinson's disease is reflected with gradual decrease of EEG delta responses during auditory discrimination. Front Psychol 2018;9:170.

[54] Moore SF, Barker RA. Predictors of Parkinson's disease dementia: towards targeted therapies for a heterogeneous disease. Parkinsonism Relat Disord 2014;20:S104–7.

[55] Gosal D, Ross OA, Toft M. Parkinson's disease: the genetics of a heterogeneous disorder. Eur J Neurol 2006;13:616–27.

[56] Backes M, Berrang P, Humbert M, Xiaoyu S, Wolf V. Simulating the large-scale erosion of genomic privacy over time. IEEE/ACM Trans Comput Biol Bioinform 2018;15:1405–12.

[57] Erlich Y, Williams JB, Glazer D, Yocum K, Farahany N, Olson M, et al. Redefining genomic privacy: trust and empowerment. PLoS Biol 2014;12:e1001983.

[58] Knoppers BM, Dove ES, Litton JE, Nietfeld JJ. Questioning the limits of genomic privacy. Am J Human Genet 2012;91:577–8, author reply 9.

[59] Korf BR. Genomic privacy in the information age. Clin Chem 2013;59:1148–50.

[60] Wang M, Ji Z, Wang S, Kim J, Yang H, Jiang X, et al. Mechanisms to protect the privacy of families when using the transmission disequilibrium test in genome-wide association studies. Bioinformatics 2017;33:3716–25.

[61] Chen F, Wang S, Jiang X, Ding S, Lu Y, Kim J, et al. PRINCESS: Privacy-protecting rare disease international network collaboration via encryption through software guard extensionS. Bioinformatics 2017;33:871–8.

[62] Baldacci F, Policardo L, Rossi S, Ulivelli M, Ramat S, Grassi E, et al. Reliability of administrative data for the identification of Parkinson's disease cohorts. Neurol Sci 2015;36:783–6.

[63] Ohno-Machado L. Sharing data for the public good and protecting individual privacy: informatics solutions to combine different goals. J Am Med Inform Assoc 2013;20:1.

[64] Lill CM, Roehr JT, McQueen MB, Kavvoura FK, Bagade S, Schjeide BM, et al. Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: the PDGene database. PLoS Genet 2012;8:e1002548.

[65] Taccioli C, Tegner J, Maselli V, Gomez-Cabrero D, Altobelli G, Emmett W, et al. ParkDB: a Parkinson's disease gene expression database. Database (Oxford) 2011;2011:bar007.

[66] Cruts M, Theuns J, Van Broeckhoven C. Locus-specific mutation databases for neurodegenerative brain diseases. Hum Mutat 2012;33:1340–4.

[67] Yang JO, Kim WY, Jeong SY, Oh JH, Jho S, Bhak J, et al. PDbase: a database of Parkinson's disease-related genes and genetic variation using substantia nigra ESTs. BMC Genomics 2009;10:S32.

[68] The Parkinson Progression Marker Initiative. The Parkinson progression marker initiative (PPMI). Prog Neurobiol 2011;95:629–35.

[69] Tang Y, Yan W, Chen J, Luo C, Kaipia A, Shen B. Identification of novel microRNA regulatory pathways associated with heterogeneous prostate cancer. BMC Syst Biol 2013;7:S6.

[70] Wang Y, Chen J, Li Q, Wang H, Liu G, Jing Q, et al. Identifying novel prostate cancer associated pathways based on integrative microarray data analysis. Comput Biol Chem 2011;35:151–8.

[71] Yang B, Li M, Tang W, Liu W, Zhang S, Chen L, et al. Dynamic network biomarker indicates pulmonary metastasis at the tipping point of hepatocellular carcinoma. Nat Commun 2018;9:678.

[72] Yu X, Zhang J, Sun S, Zhou X, Zeng T, Chen L. Individual-specific edge-network analysis for disease prediction. Nucleic Acids Res 2017;45:e170.

[73] Zhu F, Liu Q, Zhang X, Shen B. Protein interaction network constructing based on text mining and reinforcement learning with application to prostate cancer. IET Syst Biol 2015;9:106–12.

[74] Choi H, Ha S, Im HJ, Paek SH, Lee DS. Refining diagnosis of Parkinson's disease with deep learning-based interpretation of dopamine transporter imaging. Neuroimage Clin 2017;16:586–94.

[75] Shen L, Lin Y, Sun Z, Yuan X, Chen L, Shen B. Knowledge-guided bioinformatics model for identifying autism spectrum disorder diagnostic microRNA biomarkers. Sci Rep 2016;6:39663.

[76] Zhang W, Landback P, Gschwend AR, Shen B, Long M. New genes drive the evolution of gene interaction networks in the human and mouse genomes. Genome Biol 2015;16:202.

[77] Gasser T. Personalized Medicine Approaches in Parkinson's disease: the genetic perspective. J Parkinsons Dis 2016;6:699–701.

[78] Titova N, Chaudhuri KR. Personalized medicine in Parkinson's disease: time to be precise. Mov Disord 2017;32:1147–54.