

## ARTICLE OPEN



# Genome-wide association study using whole-genome sequencing identifies risk loci for Parkinson's disease in Chinese population

Hongxu Pan<sup>1,35</sup>, Zhenhua Liu<sup>1,2,35</sup>, Jinghong Ma<sup>3</sup>, Yuanyuan Li<sup>4</sup>, Yuwen Zhao<sup>1</sup>, Xiaoxia Zhou<sup>1</sup>, Yaqin Xiang<sup>1</sup>, Yige Wang<sup>1</sup>, Xun Zhou<sup>1</sup>, Runcheng He<sup>1</sup>, Yali Xie<sup>2</sup>, Qiao Zhou<sup>2</sup>, Kai Yuan<sup>2,5</sup>, Qian Xu<sup>1</sup>, Qiying Sun<sup>6</sup>, Junling Wang<sup>1</sup>, Xinxiang Yan<sup>1</sup>, Hainan Zhang<sup>7</sup>, Chunyu Wang<sup>7</sup>, Lifang Lei<sup>8</sup>, Weiguo Liu<sup>9</sup>, Xuejing Wang<sup>10</sup>, Xuebing Ding<sup>10</sup>, Tao Wang<sup>11</sup>, Zheng Xue<sup>12</sup>, Zhentao Zhang<sup>13</sup>, Ling Chen<sup>14</sup>, Qing Wang<sup>15</sup>, Yonghong Liu<sup>16</sup>, Jiayu Tang<sup>17</sup>, Xuewei Zhang<sup>18</sup>, Shifang Peng<sup>18</sup>, Chaodong Wang<sup>3</sup>, Jianqing Ding<sup>19</sup>, Chunfeng Liu<sup>20</sup>, Lijuan Wang<sup>21</sup>, Haibo Chen<sup>22</sup>, Lu Shen<sup>1</sup>, Hong Jiang<sup>1,23</sup>, Xinyin Wu<sup>24</sup>, Hongzhuan Tan<sup>24</sup>, Dan Luo<sup>25</sup>, Shuiyuan Xiao<sup>25</sup>, Xiang Chen<sup>26</sup>, Jieqiong Tan<sup>27</sup>, Zhengmao Hu<sup>27</sup>, Chao Chen<sup>27</sup>, Kun Xia<sup>27</sup>, Zhuohua Zhang<sup>5,27</sup>, Jia Nee Foo<sup>28</sup>, Cornelis Blauwendraat<sup>29</sup>, Mike A. Nalls<sup>29,30,31</sup>, Andrew B. Singleton<sup>29,30</sup>, Jun Liu<sup>4</sup>, Piu Chan<sup>3</sup>, Houfeng Zheng<sup>32</sup>, Jinchen Li<sup>1,2,27</sup>, Jifeng Guo<sup>1,2,23</sup>, Jian Yang<sup>33</sup>✉, Beisha Tang<sup>1,2,23,27</sup>✉ and the Parkinson's Disease & Movement Disorders Multicenter Database and Collaborative Network in China (PD-MDCNC)\*

Genome-wide association studies (GWASs) have identified numerous susceptibility loci for Parkinson's disease (PD), but its genetic architecture remains underexplored in populations of non-European ancestry. To identify genetic variants associated with PD in the Chinese population, we performed a GWAS using whole-genome sequencing (WGS) in 1,972 cases and 2,478 controls, and a replication study in a total of 8209 cases and 9454 controls. We identified one new risk variant rs61204179 ( $P_{\text{combined}} = 1.47 \times 10^{-9}$ ) with low allele frequency, four previously reported risk variants (*NUCKS1/RAB29*-rs11557080, *SNCA*-rs356182, *FYN*-rs997368, and *VPS13C*-rs2251086), as well as three risk variants in *LRKK2* coding region (A419V, R1628P, and G2385R) with genome-wide significance ( $P < 5 \times 10^{-8}$ ) for PD in Chinese population. Moreover, of the reported genome-wide significant risk variants found mostly in European ancestry populations, the correlation coefficient ( $r_b$ ) of effect size accounting for sampling errors was 0.91 between datasets and 63.6% attained  $P < 0.05$  in Chinese population. Accordingly, we estimated a heritability of 0.14–0.18 for PD, and a moderate genetic correlation between European ancestry and Chinese populations ( $r_g = 0.47$ ,  $se = 0.21$ ). Polygenic risk score (PRS) analysis revealed that individuals with PRS values in the highest quartile had a 3.9-fold higher risk of developing PD than the lowest quartile. In conclusion, the present GWAS identified PD-associated variants in Chinese population, as well as genetic factors shared among distant populations. Our findings shed light on the genetic homogeneity and heterogeneity of PD in different ethnic groups and suggested WGS might continue to improve our understanding of the genetic architecture of PD.

npj Parkinson's Disease (2023)9:22; <https://doi.org/10.1038/s41531-023-00456-6>

## INTRODUCTION

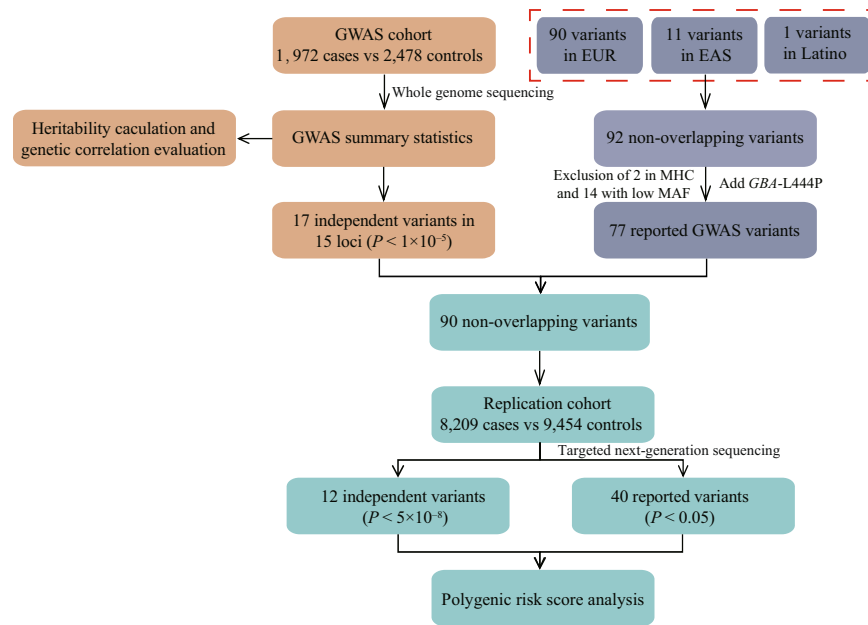
Parkinson's disease (PD) is a common age-related neurodegenerative disorder, characterised by bradykinesia, postural instability, rigidity, and resting tremors. The clinical picture includes non-motor symptoms such as cognitive impairment, autonomic dysfunction, disorders of sleep, depression and hyposmia<sup>1</sup>. The pathological hallmarks of PD are loss of dopaminergic neurons and Lewy bodies in the substantia nigra<sup>2</sup>. In China, the number of individuals with PD is estimated to reach 4.94 million by 2030, accounting for half of global cases<sup>3</sup>. The exact causes of PD are still unclear, but genetic factors, environmental factors and aging play essential roles in pathogenesis of PD<sup>4</sup>.

Although pathogenic mutations in disease-causing genes play essential roles in PD, few patients carry pathogenic mutations<sup>5,6</sup>. Common variants with small effect size can be detected via genome-wide association studies (GWASs), and have been shown to affect a large proportion of genetic predisposition of PD<sup>7</sup>. Previous GWASs of PD in European ancestry, East Asian, and Latino populations comprising more than one million samples have identified over 80 risk loci with more than 90 independent risk

variants, including in *SNCA*, *LRKK2*, *MAPT*, *BST1*, *GCH1*, *VPS13C*, and *TMEM175*<sup>8–10</sup>. Genetic heterogeneity among populations has been inferred from differences in risk loci, such as the overall genetic architecture defined by allele frequency and linkage disequilibrium (LD) patterns, and allelic heterogeneity between European ancestry and East Asian individuals<sup>8,11</sup>. GWAS data have pointed to several biological pathways involved in PD pathogenesis, such as mitochondrial function, lysosomal function, and endocytosis<sup>8,11</sup>. However, a large proportion of target subjects were of European ancestry, extrapolating these results to other populations such as Chinese becomes difficult and there is no large GWAS study of PD in Chinese population yet.

The present study explored the genetic architecture of PD with the aim of identifying risk loci responsible for its pathogenesis in the Chinese population. We conducted a GWAS in a total of 1972 PD cases and 2478 controls using whole-genome sequencing (WGS), as well as a replication study in 8209 cases and 9454 controls using multiplex PCR sequencing. We also compared genome-wide significant variants for PD, reported mostly in population of European ancestry, with our results of Chinese population. Finally, we estimated the

A full list of author affiliations appears at the end of the paper.



**Fig. 1 Workflow of the study.** EUR European, EAS East Asian, MHC histocompatibility complex region, MAF minor allele frequency.

heritability of PD, assessed its genetic correlation with other traits, and established a predictive model to evaluate at-risk individuals in the Chinese population.

## RESULTS

### Cohort description and quality control

A total of 22,366 subjects were included in this study: 1980 patients with PD and 2516 matched controls in the GWAS cohort, plus 8294 cases and 9576 matched controls in the replication cohort. Details about genotype quality of GWAS and replication cohorts are provided in Supplementary Table 1 and Supplementary Table 2. The final subjects included in the statistical analysis after quality control (Supplementary Table 3) encompassed 1972 cases (the mean age at recruitment,  $66.76 \pm 7.08$  years; the mean age at onset,  $61.88 \pm 6.93$  years) and 2478 controls (the mean age at recruitment,  $62.32 \pm 7.11$  years) in the GWAS cohort, and 8209 cases (the mean age at recruitment,  $60.23 \pm 11.20$  years; the mean age at onset,  $57.77 \pm 11.95$  years) and 9454 controls (the mean age at recruitment,  $64.29 \pm 9.68$  years) in the replication cohort. Detailed demographic information is listed in Supplementary Table 4.

To explore the quality of WGS with the medium degree of depth, 54 samples from the GWAS cohort were randomly picked using Excel and subjected to WGS with a high degree of depth setting, whereas another randomly picked 39 samples were re-sequenced with the same WGS as GWAS cohort. Sequencing results were compared within sample pairs, revealing good genotype concordance between the analysed GWAS variants (Supplementary Fig. 1). The workflow of this study was showed in Fig. 1.

### Genome-wide association analysis

In the GWAS cohort, 6,910,086 autosomal single nucleotide polymorphisms (SNPs) passed the quality control threshold and were included in further analyses. Principal component analysis confirmed a good match between patients with PD and controls (Fig. 2a, b). A quantile-quantile plot indicated that population stratification had negligible effects on the statistical results ( $\lambda$  genomic control = 1.065, LD score intercept = 1.036; (Fig. 2c). Two known risk loci (*NUCKS1/RAB29* and *SNCA*) exhibited

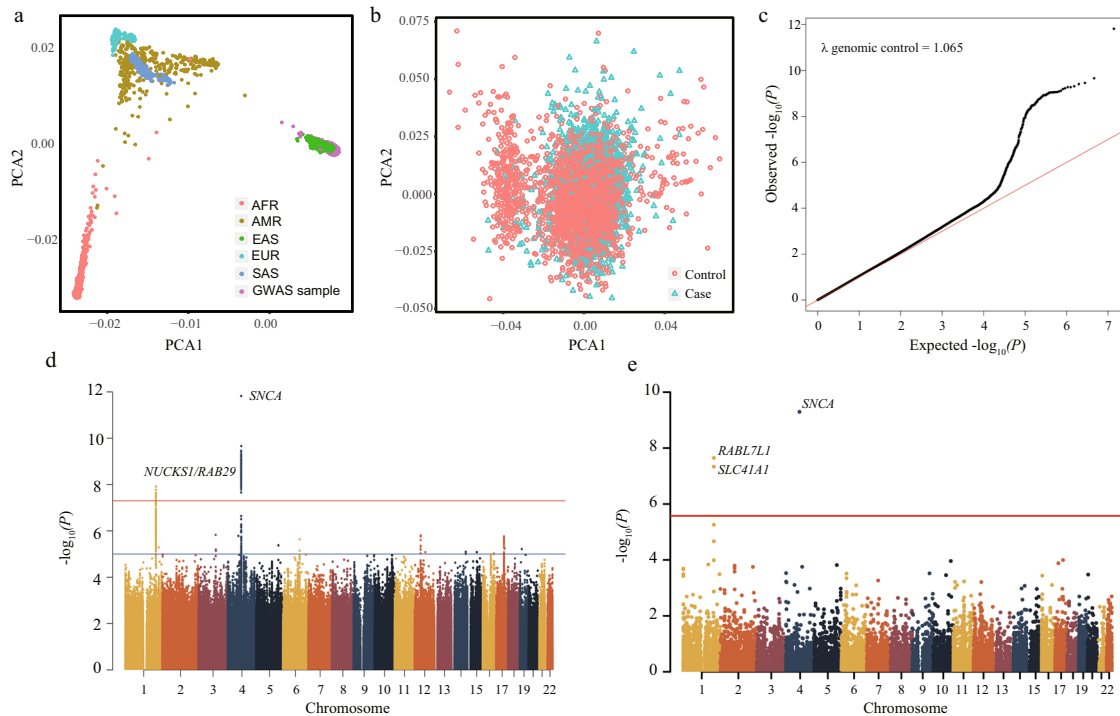
genome-wide significance ( $P < 5 \times 10^{-8}$ ), and 13 risk loci exhibited suggestive significance ( $P < 1 \times 10^{-5}$ ) in the GWAS analysis. Of the 13 suggestive risk loci, three were known (*LRRK2*, *FYN*, and *VPS13C*) and 10 were possibly novel (Fig. 2d; Supplementary Table 5). *NUCKS1/RAB29*, *SNCA*, and *LRRK2* regions were previously reported to harbour multiple independent signals, and conditional analysis of our GWAS data revealed probable multiple signals in *LRRK2*, but not in *NUCKS1/RAB29* and *SNCA* regions (Supplementary Table 6). In the *LRRK2* locus, three previously reported risk variants, rs34594498 (*LRRK2*-A419V), rs33949390 (*LRRK2*-R1628P), and rs34778348 (*LRRK2*-G2385R), were found to be independently associated with PD with suggestive significance ( $P < 1 \times 10^{-5}$ ) (Fig. 3).

In the replication cohort, 17 independent risk variants in 15 risk loci were genotyped and analysed. In the combined analysis, eight of these SNPs showed associations exceeding the genome-wide significance threshold ( $P < 5.0 \times 10^{-8}$ ; Table 1), including four reported GWAS SNPs, three risk variants in the *LRRK2* gene and one newly identified risk variant rs61204179.

The new risk variant (rs61204179) was located in the ninth intron of the *HEATR6* gene. This variant had a minor allele frequency of 4% in our population and low LD with other variants in the *HEATR6* region ( $r^2 < 0.2$  with most variants in the East Asian individuals of 1000 Genomes Project data, Fig. 4, Supplementary Fig. 2). Notably, rs61204179 was rare in the populations of European ancestry, making it a Chinese-specific PD risk variant, similar to the associated variants (A419V, R1628P and G2385R) in *LRRK2* gene. *HEATR6* shows high expression in multiple brain regions including midbrain (Supplementary Fig. 3) and was reported to be downregulated in the peripheral blood of patients with PD<sup>12</sup>.

### Gene-based association analysis

Besides the GWAS of common SNPs, we conducted genome-wide gene association analysis using GWAS data by MAGMA. MAGMA relies on converging evidence from multiple genetic variants in the same gene and can yield novel signals at a gene-based level. Here, it identified three associated genes (*RAB7L1*, *SLC41A1*, and *SNCA*), all of which were located in genome-wide significant loci by the GWAS (Fig. 2e; Supplementary Table 7).



**Fig. 2** Association results of GWAS cohort. **a** PCA plot of samples projecting to the 1000 Genomes subjects. **b** PCA of our 1972 PD cases and 2478 controls in GWAS cohort. **c** The quantile–quantile (Q–Q) plot of the association results from the GWAS cohort. **d** Manhattan plot of  $P$  values on the  $-\log_{10}$  scale for 6,910,866 SNPs in the GWAS cohort. **e** Manhattan plot of  $P$  values on the  $-\log_{10}$  scale for genome-wide gene-based analysis in the GWAS stage.

### Validation of reported GWAS variants

To date, 78 loci and 90 independent variants have been reported in populations of European ancestry, 11 loci (11 variants) in the East Asian population, and one locus (one variant) in the Latino population<sup>8–10</sup>. Among the 92 independent PD-associated variants in the 80 non-overlapping loci, a high correlation of minor allele frequency was observed between the East Asian and European ancestry populations ( $r = 0.65$ , Fig. 5a, b). Fourteen variants with minor allele frequency  $< 0.5\%$  in the East Asian population and two variants located in the major histocompatibility complex region, were excluded from the validation analysis. Besides, the *GBA*-L444P variant was added in validation given the reported genome-wide significant risk variant in *GBA* gene (N370S) in European ancestry populations was very rare in the Chinese population but *GBA*-L444P was reported to be associated with PD in Chinese.

Of the 77 reported variants included in our validation analysis, nine attained genome-wide significance in the combined data, 15 were significant after Bonferroni correction ( $P < 0.05/77$ ), and 25 had  $P < 0.05$  (a total of 63.6% variants were validated, Fisher's exact test  $P < 2.84 \times 10^{-15}$ ). All of the validated variants had the same effect direction as the reported, and of the 28 non-validated variants, 24 exhibited the same effect direction as reported previously (Supplementary Table 8), suggesting that more samples were needed to further validate these variants (sign concordance test  $P < 1.99 \times 10^{-9}$ ). Specifically, rs11557080 and rs823118 were reported to be independent risk variants in the *NUCKS1/RAB29* locus in populations of European ancestry, but they appeared in high LD in our data ( $r^2 = 0.94$ ), and conditional analysis showed that only one signal remained significant (Supplementary Table 8). Moreover, the estimation of variants effect correlation ( $r$ ) between our data and existing data<sup>8,9</sup> was 0.82, and the correlation estimate ( $r_b$ ) was 0.91 when accounting for errors in the variant's effects (Fig. 5c, d), suggesting a large proportion of GWAS markers discovered in Europeans were likely replicable in the Chinese

population. Nevertheless, some differences between the associated sites still persist.

### Heritability, genetic correlation and MR

Using the genetic data of the GWAS cohort, the estimated heritability of PD in the Chinese population was 0.140 (se = 0.047) by LDSC and 0.184 (se = 0.041) by GCTA-GREML, assuming a global prevalence of 0.5%. These values seemed lower than the estimate of 0.22 for populations of European ancestry<sup>8</sup>, but the heritability remained comparable between two populations given the relatively large confidence interval of heritability estimates.

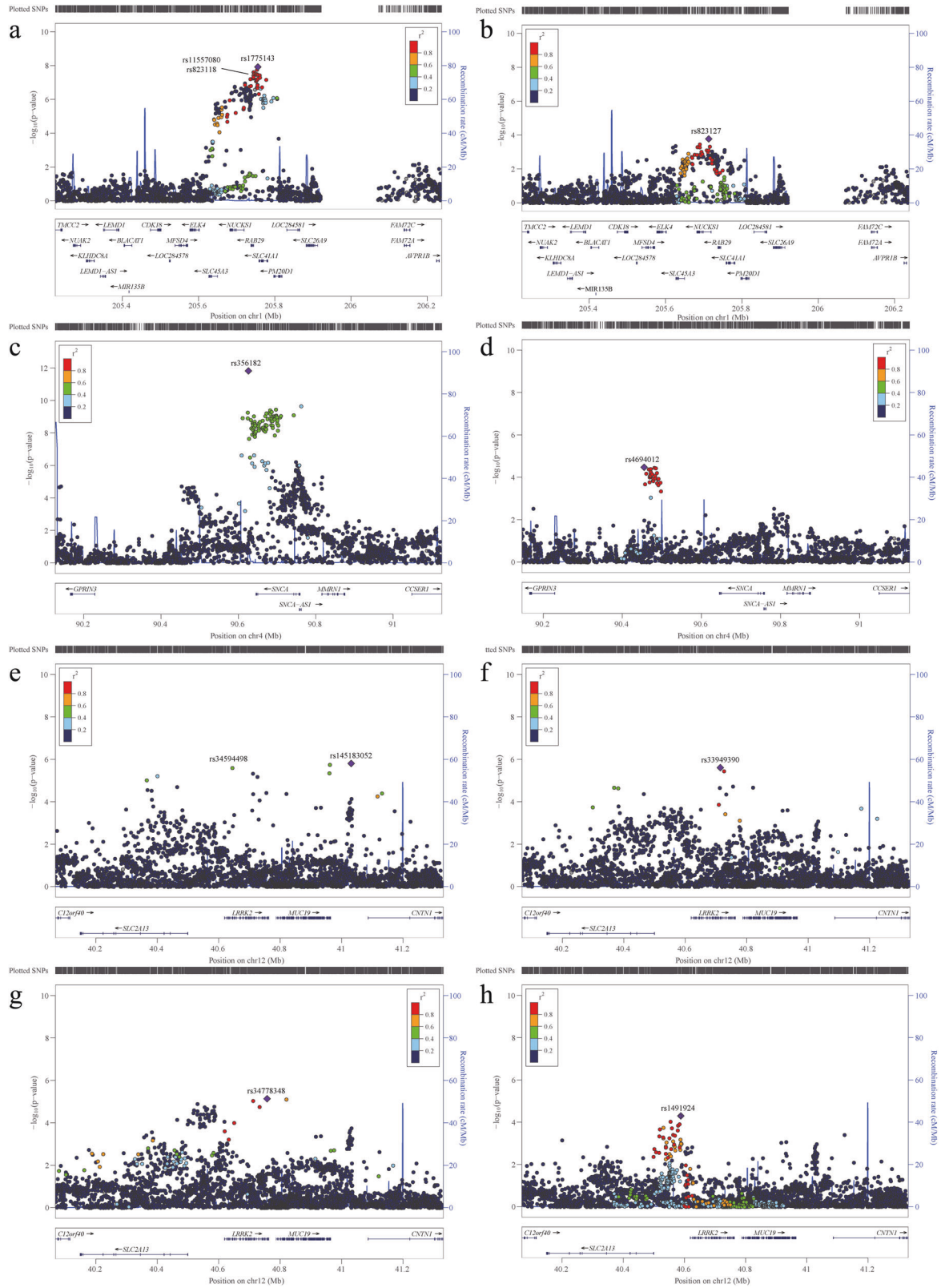
Using our GWAS data and publicly available PD-GWAS summary statistics, we found that the European ancestry and Chinese populations showed an intermediate genetic correlation ( $r_g = 0.47$ , se = 0.21), suggesting a certain amount of population heterogeneity in PD. This result was in line with the validation outcomes of GWAS variants and further confirmed the need for genome research on PD among different populations.

We further analysed possible genetic correlations between Chinese patients with PD and other phenotypes in East Asian populations. Using publicly available GWAS summary data originating mainly from the BioBank Japan Project, 157 traits were analysed, but no significant genetic correlation was found after Bonferroni correction. However, height, type 2 diabetes, rheumatoid arthritis, systemic lupus erythematosus, and gamma-glutamyl transferase showed suggestive association with PD ( $P < 0.05$ ) (Supplementary Table 9).

Additional MR analysis between PD and phenotypes with suggestive genetic correlation refuted any significant causal relationship ( $P > 0.05$ ) (Supplementary Table 10).

### Polygenic risk score analysis

Using the GWAS cohort as a training dataset for effect size calculation and the validation cohort as test data, we calculated



**Fig. 3 Regional plot of conditional analysis of reported loci with multiple independent signals.** **a, b** Conditional analysis in *NUCKS1/RAB29* region. **c, d** Conditional analysis in *SNCA* region. **e–h** Conditional analysis in *LRRK2* region. Reference data of Asian population from 1000 Genomes were used for linkage disequilibrium calculation.



**Table 1.** Association results of eight genome-wide significant variants in combined analysis.

| Tag SNP    | Position (hg19) | Nearest gene | Effect allele | Other allele | MAF_EAS | MAF_EUR | GWAS cohort |       | Replication cohort |        |       | Combined |        |       | $P_{het}$ |       |
|------------|-----------------|--------------|---------------|--------------|---------|---------|-------------|-------|--------------------|--------|-------|----------|--------|-------|-----------|-------|
|            |                 |              |               |              |         |         | Beta        | se    | Beta               | se     | P     | Beta     | se     | P     |           |       |
| rs11557080 | 1:205737739     | NUCKS1/RAB29 | G             | A            | 0.457   | 0.855   | -0.249      | 0.046 | 6.08E-08           | -0.206 | 0.024 | 1.80E-18 | -0.215 | 0.021 | 1.04E-24  | 0.399 |
| rs356182   | 4:90626111      | SNCA         | A             | G            | 0.332   | 0.650   | -0.344      | 0.049 | 1.52E-12           | -0.342 | 0.025 | 2.56E-42 | -0.343 | 0.022 | 3.09E-53  | 0.970 |
| rs997368   | 6:112243291     | FYN          | G             | A            | 0.341   | 0.128   | -0.160      | 0.046 | 5.37E-04           | -0.109 | 0.023 | 3.39E-06 | -0.119 | 0.021 | 1.11E-08  | 0.326 |
| rs34594498 | 12:40646786     | LRRK2        | T             | C            | 0.007   | 0       | 0.984       | 0.209 | 2.47E-06           | 0.855  | 0.108 | 2.64E-15 | 0.883  | 0.096 | 4.05E-20  | 0.586 |
| rs33949390 | 12:40713845     | LRRK2        | A             | G            | 0.028   | 0.0006  | 0.661       | 0.144 | 4.57E-06           | 0.503  | 0.076 | 4.02E-11 | 0.544  | 0.067 | 6.68E-16  | 0.343 |
| rs34778348 | 12:40757328     | LRRK2        | C             | G            | 0.029   | 0       | 0.452       | 0.110 | 3.71E-05           | 0.643  | 0.061 | 4.26E-26 | 0.608  | 0.053 | 3.22E-30  | 0.233 |
| rs2251086  | 15:61997385     | VP513C       | T             | C            | 0.162   | 0.149   | -0.191      | 0.061 | 1.77E-03           | -0.158 | 0.030 | 1.56E-07 | -0.164 | 0.027 | 1.19E-09  | 0.622 |
| rs61204179 | 17:58139282     | HEATR6       | T             | C            | 0.043   | 0       | 0.477       | 0.105 | 5.86E-06           | 0.242  | 0.054 | 7.91E-06 | 0.291  | 0.048 | 1.47E-09  | 0.047 |

MAF data were extracted from gnomAD database.

Beta regression coefficient, se standard error of beta,  $P_{het}$  the P-value of heterogeneity, MAF minor allele frequency, EUR European, EAS East Asian.

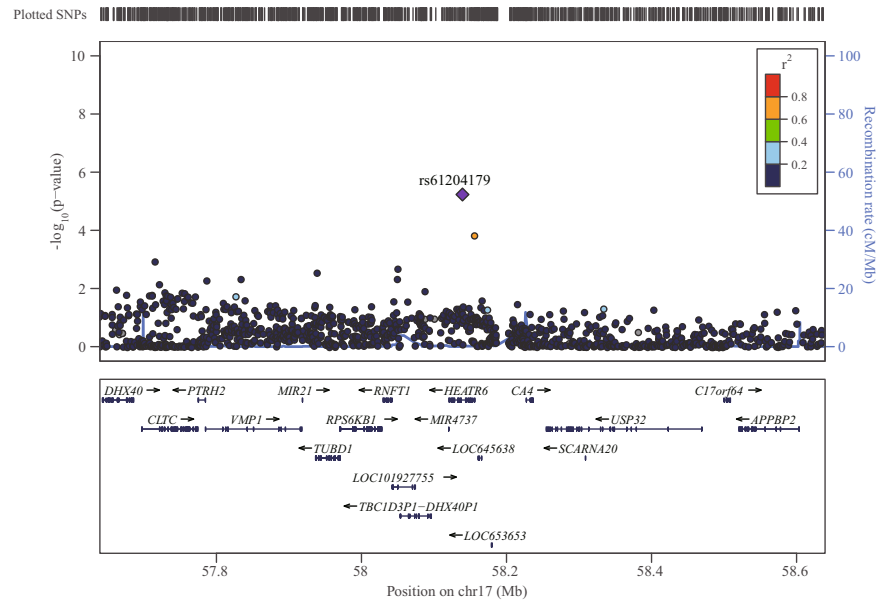
polygenic risk scores based on three sets of SNPs stratified by significance level. For 12 genome-wide significant variants, the proportion of explained genetic liability of PD was 16% based on our current heritability estimates. Moreover, 12 genome-wide significant variants ( $P < 5 \times 10^{-8}$ ) and 15 variants with ( $P < 0.05/77$ ) explained 20% of PD heritability, while 52 variants with  $P < 0.05$  accounted for 23% of heritability (Fig. 6a). In the weighted polygenic risk score distribution based on different single nucleotide variant sets, we observed a 2.8-fold and 3.9-fold difference in risk between the top and bottom 25% of the test population (Fig. 6b).

## DISCUSSION

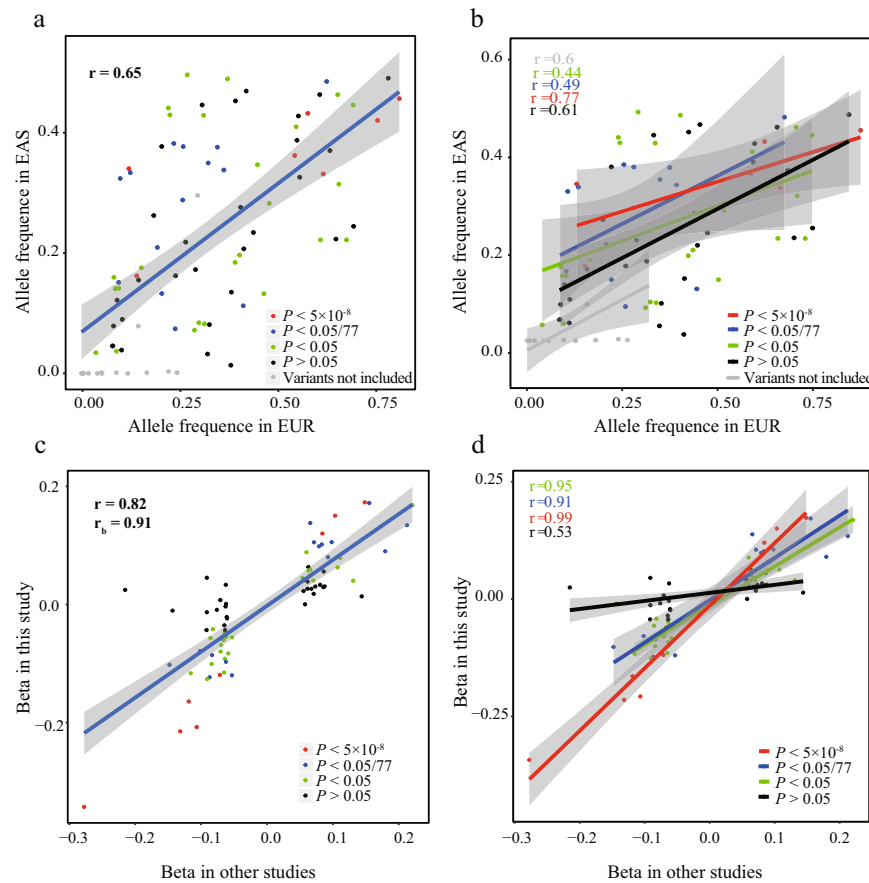
We conducted a GWAS and a replication study in a large Chinese cohort, identified a new PD-related locus, and confirmed multiple independent risk variants with genome-wide significance in the *LRRK2* locus. In addition, a large proportion of reported GWAS variants were validated in our data. We constructed polygenic risk models to better predict the risk of PD and also systematically assessed the genetic correlation of PD in different populations, and in relation to other phenotypes among East Asians. Overall, these analyses may provide new insights on the impact of genetic variants on the pathogenesis of PD.

SNP arrays and WGS have been shown similarly efficient at surveying common variants in the genome<sup>13</sup>. Here, we employed the WGS method for genotyping our GWAS cohort, because it captured significant signals with low minor allele frequency, which were not well imputed in commonly used SNP arrays<sup>8,9</sup>. This was the case of the newly identified risk locus in the *HEATR6* gene and multiple independent risk variants located in the *LRRK2* region, supporting the potential use of WGS for future GWASs. On the other hand, the newly identified locus in the *HEATR6* gene could not be validated in the published PD GWAS datasets of Asian populations<sup>9,14,15</sup> due to the methodology differences. In addition, the tag SNP of the new risk locus was a rare variant in populations of European ancestry and, therefore, it could not be validated in PD GWAS datasets of European ancestry populations either<sup>8,15</sup>. Given that most currently used functional annotation databases are derived from populations of European ancestry<sup>16–18</sup>, while the corresponding databases for East Asian populations are still too narrow in scope<sup>19</sup>, it is hard to comprehensively explore the possible regulatory mechanism of this locus. Based on data from European ancestry populations<sup>20–22</sup>, there is few genomic regulatory elements in the region harbouring this locus. *HEATR6* encodes the HEAT repeat containing protein 6, which have 37–47 amino acid motifs called HEAT repeats forming repetitive arrays of short amphiphilic  $\alpha$ -helices. While the functions of the *HEATR6* remains unclear, HEAT repeat proteins mediate important protein-protein interactions involved in cytoplasmic and nuclear transport, microtubule dynamics and chromosome segregation<sup>23</sup>. Several HEAT repeat proteins were reported to cause neurological diseases<sup>23–25</sup>. *HEATR6* showed high expression in multiple brain regions including midbrain, moreover, *Heatr6* were reported to be altered in neurological diseases<sup>26,27</sup> and downregulated in the peripheral blood of patients with PD<sup>12</sup>. We speculate the new risk variant rs61204179 may mediate the risk of PD through *HEATR6*, but functional studies are required to further confirm the relationship between this locus and PD.

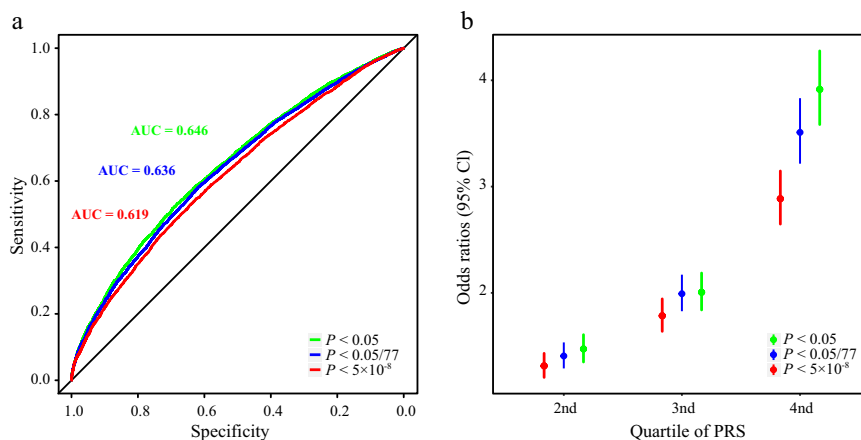
Multiple independent risk variants were identified by GWASs in the *LRRK2* region of European ancestry populations<sup>8,28</sup>, and this observation was confirmed in Chinese with *LRRK2*-A419V, *LRRK2*-G2385R, and *LRRK2*-R1628P showing associations independently. However, the risk variants of *LRRK2* exhibited significantly different incidence across populations. For example, the *LRRK2*-G2019S which was identified in European ancestry populations is very rare in the Chinese population; whereas multiple independent variants identified in the present study are rare in the



**Fig. 4 Regional plot of the rs61204179.** Regional plot of the association signal of rs61204179 by LocusZoom. Reference data of Asian population from 1000 Genomes were used for linkage disequilibrium calculation.



**Fig. 5 Correlation plots.** **a, b** Allele frequency—Allele frequency plot showing the correlation of minor allele frequencies for reported genome-wide significant variants between European (x-axis) and East Asian populations (y-axis); allele frequencies were extracted from gnomAD database. **c, d** Beta-beta plot showing the correlation between effect sizes derived from the GWAS in which they were reported (x-axis) and effect sizes calculated in this study (y-axis); beta in other studies were from European population (Nalls, 2019) and East Asian population (Foo, 2020). Red dots denote variants with  $P < 5 \times 10^{-8}$ , blue dots denote variants with  $P < 0.05/77$ , green dots denote variants with  $P < 0.05$ , black dots denote variants with  $P > 0.05$ , and grey dots denote variants not included in replication cohort of this study.  $r$  shows the Pearson correlation coefficient and  $r_b$  shows the correlation coefficient correcting for errors in the estimated effects. EUR European, EAS East Asian.



**Fig. 6 Predictive model.** **a** Polygenic risk score (PRS) receiver-operator curves for variants with different significance levels. **b** Odds ratio of developing PD for each quartile of the PRS compared with the lowest quartile of genetic risk. Red lines and dots correspond to data calculated from variants with  $P < 5 \times 10^{-8}$ , blue lines and dots correspond to data calculated from variants with  $P < 0.05/77$ , and green lines and dots correspond to data calculated from variants with  $P < 0.05$ .

European ancestry populations<sup>7,8</sup>. In addition, although the reported GWAS signal for rs76904798, located in the 5' untranslated region of *LRRK2*, was common in both populations, no significant association was found in the Chinese population, indicating substantial population heterogeneity in the *LRRK2* locus. This finding is of relevance for future studies, which should further explore PD-related variants in the *LRRK2* region, such as finding more disease-associated variants and fine mapping of the causal variants across different populations.

We systematically validated the reported GWAS variants and found that those identified in populations of European ancestry were well validated in our population. In fact, the correlation of effect size between European ancestry and East Asian populations was significantly better than the previously published correlation (0.82 vs. 0.44)<sup>9</sup>. The correlation was even higher (0.91) after applying  $r_b$  method which estimate the correlation of estimated SNP effects accounting for estimation errors. Moreover, after the stratification of variant's significance, the effects of validated variants were better correlated between datasets. This result pointed to the genetic homogeneity of PD between the two populations and indirectly confirmed the success of our methodology. One of the two risk variants identified in Eastern Asia have not been validated in our study (SV2C-rs246814,  $P = 0.49$ ), despite the strong association of SV2C-rs246814 was observed in further replication study in both European ancestry population and pooling dataset of worldwide population<sup>29</sup>. Another risk variant was validated in our data (WBSR17-rs9638616,  $P = 0.047$ ), which was consistent with former studies<sup>9,29</sup>. Future studies are required to investigate PD among populations inhabiting different regions of East Asia.

No significant genetic correlation was found between traits describing the East Asian population and the PD in Chinese population, however, suggestively significant correlated traits found in this study (including type 2 diabetes, autoimmune disorders and height) may be informative for further epidemiological, clinical, and pathophysiological researches<sup>30</sup>. The PRS model was constructed in this study, but the overall performance of PD prediction was limited, and there was no significant improvement compared with the previous studies.

This study has some limitations. First, although the total sample size exceeded 20,000, this study had moderate power because the subjects were divided into GWAS and validation cohorts, resulting in limited power to detect variants with small or moderate effects. This may be the reason for the limited number of identified new loci and the high false-positive rate of nominally significant GWAS signals. Second, we only validated tag SNPs in the associated loci,

but not in the entire locus, which could result in the loss of valuable information for fine-mapping based on different LD structures within populations. Third, the newly identified risk signal has not been verified by functional studies, and requires further validation in larger cohorts.

In conclusion, this GWAS using WGS identified PD risk loci in Chinese population and validated shared genetic factors among PD from different populations. These findings further emphasise both the genetic homogeneity and heterogeneity of PD in disparate populations, and suggested the potential of WGS in improving our understanding of the disease's genetic architecture.

## METHODS

### Participants

A total of 10,274 patients with PD and 12,092 control subjects free from neurological disorders were recruited from study groups participating in the Parkinson's Disease and Movement Disorders Multicenter Database and Collaborative Network in China (<http://pd-mdcnc.com>)<sup>31</sup> and centres across China. The GWAS included 1980 cases with PD and 2516 matched controls due to the costs of WGS and the numbers of study participants recruited during the GWAS phase. Another 8,294 cases and 9,576 controls were included in the replication study. Each patient was diagnosed by two neurologists specialised in movement disorders according to either the Movement Disorder Society Clinical Diagnostic Criteria<sup>32</sup> or the United Kingdom Parkinson's Disease Society Brain Bank Clinical Diagnostic Criteria<sup>33</sup> in recruiting sites. Blood samples were collected, and genomic DNA was prepared according to standard procedures. The study was approved by the ethics committees of Xiangya Hospital of Central South University and other centres including Xuanwu Hospital of Capital Medical University, Ruijin Hospital of Shanghai Jiao Tong University School of Medicine, Affiliated Brain Hospital of Nanjing Medical University, Union Hospital of Huazhong University of Science and Technology, the First Affiliated Hospital of Zhengzhou University and First Affiliated Hospital of Sun Yat-sen University. Written informed consent was obtained from the participants or their legal guardians.

### Next-generation sequencing and variants calling

For the GWAS, WGS was performed on the Illumina NovaSeq 6000 sequencing platform with 150-bp pair-end reads, followed by preparation of an Illumina DNA library using genomic DNA

sheared to 100–1000 bp. The mean WGS depth was 13× per individual in the GWAS cohort.

For the replication study, a single multiplex PCR was used to amplify the amplicons of candidate variants from our GWAS, as well as variants reported in previous GWASs on European ancestry, East Asian, and Latino populations<sup>8–10</sup>, followed by sequencing on an Illumina platform. The mean depth was 2000× per individual in the replication cohort.

Variant calling was conducted on all samples using BWA<sup>34</sup> and GATK4<sup>35</sup>. Clean reads were aligned to the GRCh37 human reference genome (hg19) using the BWA mem tool to produce SAM files. The SAMtools<sup>36</sup> view option was used to convert SAM files into BAM files. MarkDuplicates was used to mark PCR duplicates with a REMOVE\_DUPLICATES parameter setting of false. The BaseRecalibrator generated a recalibration table using known dbSNP and 1000 Genomes Project VCF resources. ApplyBQSR generated the recalibrated final BAM files for HaplotypeCaller. The gVCF file for each sample was obtained and combined into joint VCF files using GATK4 GenomicsDBImport and GenotypeGVCFs, following the suggested pipelines. We calculated the VQSLOD value of each single nucleotide variant (SNV) and insertion-deletion (INDEL) variant by setting the max-Gaussian value to 5. SNVs were annotated with ReadPosRankSum, MQRankSum, DP, QD, FS, and SOR; whereas INDELS were annotated with ReadPosRankSum, DP, QD, and FS. In the ApplyVQSR filtering step, the true sensitivity level for INDELS and SNVs was set to 99.0 and 99.6, respectively. All variants that passed the filter were retained for downstream analyses. We conducted LD-based genotype refinement for low-confidence genotypes and missing sites in WGS data using BEAGLE v5.1<sup>37</sup> with default settings.

### Quality control

For individual and variant quality control of WGS data, the data were recoded into binary PLINK input format using PLINK v1.9<sup>38</sup>. Variant quality control was accomplished by keeping all autosomal SNPs with minor allele frequency  $\geq 0.01$  which may have sufficient power to detect the associations given the sample size in GWAS cohort<sup>13</sup>, while removing variants deviating from the Hardy-Weinberg equilibrium ( $P < 1 \times 10^{-4}$ ). Individuals and variants that passed the quality control thresholds were subjected to further analyses. Individuals were excluded if they showed conflicting sex assignments between inferred sex and self-reported sex, deviating heterozygosity/genotype calls ( $\pm 3$  standard deviations), or cryptic relatedness (identity by descent  $> 0.15$ ). The remaining samples were assessed for population outliers and stratification by principal component analysis using PLINK v1.9, and those that showed divergent ancestry were excluded. For the replication cohort, variants with low-quality genotypes (Phred-scaled genotype quality score 30), low call rates (missing rate  $> 5\%$ ), or departure from Hardy-Weinberg equilibrium ( $P < 1 \times 10^{-4}$ ) were excluded. Samples with low genotype call rates (missing rate  $> 5\%$ ) were removed from further analysis. We also did principal component analysis using genotyped variants not associating with PD and excluded the outliers (Supplementary Fig. 4).

We performed WGS with a high degree of depth setting (30× depth) on 54 GWAS samples which were randomly picked using Excel from the entire list of control subjects to estimate the calling accuracy of WGS with medium depth (13× depth). Moreover, WGS with medium depth (13×) was performed again on 39 randomly picked GWAS control samples to estimate calling consistency. We computed calling concordance between matched samples for genotypes of all GWAS-analysed variants and non-reference genotypes of GWAS variants using previously specified formulae<sup>39</sup>.

### Association testing and combined analysis

For GWAS, logistic regression analysis was conducted in PLINK v1.9 to test the differences in allele dosage between patients with PD and controls. An additive genetic model adjusted for sex, age, and population substructure based on the first five principal components was employed. Conditional analyses were conducted by adding in-locus tag SNP, which was the most significant variant in our data or reported variant in linkage with the most significant variant, to logistic regression as covariates until no variants within the locus reached suggestive significance ( $P < 1 \times 10^{-5}$ ). SNPs reached suggestive significance within the corresponding locus in the GWAS, and previously reported variants in GWASs were included for validation. In the replication study, logistic regression analyses in PLINK v1.9 were conducted on genotype dosages with sex, age and first two principal components as covariates.

To improve the statistical power of the validated variants, we used meta-analysis to combine the association results from the GWAS and replication study using METAL<sup>40</sup> with an inverse variance-based model. The extent of heterogeneity was assessed using  $I^2$  and  $P$ -values of the  $Q$  statistics calculated in METAL<sup>41</sup>.

### Gene-based association analysis

SNP-based  $P$ -values from the GWAS were used as inputs for gene-based analysis of common variants. We used all 19,427 protein-coding genes from NCBI 37.3 (hg19) gene definitions as the basis for a genome-wide gene association analysis in MAGMA<sup>42</sup>. The MAGMA approach is based on a multiple linear principal component regression model. By projecting the multivariate LD matrix of SNPs in a gene estimated from the 1000 Genomes Phase 3 Asian reference population<sup>43</sup>, the principal components explaining genetic variation were extracted first. These principal components were used as predictors of PD under a linear regression framework. MAGMA then employed Fisher's test to compute  $P$ -values for association testing. A stringent Bonferroni correction was applied to account for multiple testing.

### SNP-based heritability analysis

We used linkage disequilibrium score regression (LDSC)<sup>44</sup> and univariate genome-wide complex trait analysis- genomic-relatedness-based restricted maximum-likelihood (GCTA-GREML)<sup>45</sup> with default parameters and the same covariates as GWAS to estimate the heritability of PD. East Asian LD scores were downloaded from <https://github.com/bulik/ldsc>. The LDSC method derives the heritability of PD by regressing an SNP's association statistic onto its LD score (the sum of squared correlations between the minor allele count of the SNP and the minor allele count of every other SNP). GCTA-GREML methods estimate heritability by calculating the proportion of variation in phenotypes explained by common SNPs using individual phenotypic and genotypic information.

### Genetic correlation and variant effect correlation

POPCORN<sup>46</sup> was used to investigate the genetic correlation of PD between populations of European ancestry and East Asians. POPCORN uses a Bayesian approach, which assumes that genotypes are drawn separately from each population and that effect sizes follow the infinitesimal model. Genetic correlations in POPCORN were computed in the 'genetic effect' mode, which estimated the correlation based on LD covariance scores and effect sizes from summary statistics. We used bivariate LDSC to investigate the genetic correlations of PD with multiple traits and diseases in East Asian populations based on data from the BioBank Japan Project (<http://jenger.riken.jp/en/result>) and the GWAS catalogue<sup>47</sup>. Pearson correlation analysis and  $r_b$  method<sup>48</sup> which correct sampling errors in the estimated SNP effects were



used to estimate the correlation of SNP effects between reported GWASs and our data.

### Mendelian randomisation

Mendelian randomisation (MR) was used to assess the potential causal role of traits showing significant genetic correlations with PD. A two-sample MR approach was applied using genetic effect estimates for SNPs presenting genome-wide significance for PD, and those for the risk of associated traits. The analysis was performed using the MendelianRandomization package<sup>49</sup> in R based on two different strategies. According to the inverse variance weighted (IVW) MR method, which assumed that all SNPs were valid instrumental variables, conventional linear regression analysis of Wald ratios for each SNP was undertaken and weighted by the estimated inverse variance. This method constrains the regression when  $\beta_{ZX}$  and  $\beta_{ZY}$  are equal to zero. According to the MR-Egger regression, horizontal pleiotropy was tested to provide a causal estimate in the presence of pleiotropy. As in the IVW method,  $\beta_{ZY}$  is plotted against  $\beta_{ZX}$ . However, the intercept is not fixed; therefore, the deviation from the origin provides evidence for pleiotropic effects in the corresponding direction. In the absence of unbalanced pleiotropy, if SNPs are viewed individually,  $\beta_{IV}$  values are distributed symmetrically around the point estimate, as demonstrated by a funnel plot.

### Polygenic risk prediction

Predictive modelling was performed using the polygenic risk score in the replication samples. Weighted polygenic risk scores were calculated based on SNP significance in the combined analysis and their effect size in the GWAS. Areas under the receiver operating characteristic curve were calculated by comparing the observed case/control status and the polygenic risk score calculated using PRSice2<sup>50</sup> profiling in a standard weighted allele-dose manner. Equations from Wray and colleagues<sup>51</sup> were used to estimate the proportion of genetic liability accounted by SNP sets assuming a global prevalence of 0.5%.

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### DATA AVAILABILITY

The summary data of this GWAS can be accessed after an approved application to the Open Archive for Miscellaneous Data (OMIX) of National Genomic Data Center (NGDC). The accession code is [OMIX002795](https://www.ngdc.org.cn/omic/002795).

### CODE AVAILABILITY

The related codes and scripts for the study will be made available upon reasonable request to the corresponding author.

Received: 9 September 2022; Accepted: 12 January 2023;

Published online: 09 February 2023

### REFERENCES

- Bloem, B. R., Okun, M. S. & Klein, C. Parkinson's disease. *Lancet* **397**, 2284–2303 (2021).
- Schneider, S. A. & Obeso, J. A. Clinical and pathological features of Parkinson's disease. *Curr. Top. Behav. Neurosci.* **22**, 205–220 (2015).
- Dorsey, E. R. et al. Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. *Neurology* **68**, 384–386 (2007).
- Johnson, M. E., Stecher, B., Labrie, V., Brundin, L. & Brundin, P. Triggers, facilitators, and aggravators: redefining Parkinson's disease pathogenesis. *Trends Neurosci.* **42**, 4–13 (2019).
- Zhao, Y. et al. The role of genetics in Parkinson's disease: a large cohort study in Chinese mainland population. *Brain* **143**, 2220–2234 (2020).
- Liu, Z. et al. Deficiency in endocannabinoid synthase DAGLB contributes to early onset Parkinsonism and murine nigral dopaminergic neuron dysfunction. *Nat. Commun.* **13**, 3490 (2022).
- Blauwendraat, C., Nalls, M. A. & Singleton, A. B. The genetic architecture of Parkinson's disease. *Lancet Neurol.* **19**, 170–178 (2020).
- Nalls, M. A. et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol.* **18**, 1091–1102 (2019).
- Foo, J. N. et al. Identification of risk loci for Parkinson disease in Asians and comparison of risk between Asians and Europeans: a genome-wide association study. *JAMA Neurol.* **77**, 746–754 (2020).
- Loesch, D. P. et al. Characterizing the genetic architecture of Parkinson's disease in Latinos. *Ann. Neurol.* **90**, 353–365 (2021).
- Kia, D. A. et al. Identification of Candidate Parkinson Disease Genes by Integrating Genome-Wide Association Study, Expression, and Epigenetic Data Sets. *JAMA Neurol.* **78**, 464–472 (2021).
- Tan, C., Liu, X. & Chen, J. Microarray analysis of the molecular mechanism involved in Parkinson's disease. *Parkinsons Dis.* **2018**, 1590465 (2018).
- Visscher, P. M. et al. 10 years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* **101**, 5–22 (2017).
- Satake, W. et al. Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nat. Genet.* **41**, 1303–1307 (2009).
- Kim, J. J. et al. Multi-ancestry genome-wide meta-analysis in Parkinson's disease. *medRxiv* <https://doi.org/10.1101/2022.08.04.22278432> (2022).
- Consortium, G. T. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
- Vosa, U. et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet.* **53**, 1300–1310 (2021).
- Ramasamy, A. et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat. Neurosci.* **17**, 1418–1428 (2014).
- Ota, M. et al. Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases. *Cell* **184**, 3006–3021.e3017 (2021).
- Boix, C. A., James, B. T., Park, Y. P., Meuleman, W. & Kellis, M. Regulatory genomic circuitry of human disease loci by integrative epigenomics. *Nature* **590**, 300–307 (2021).
- Consortium, E. P. et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).
- Zhang, K. et al. A single-cell atlas of chromatin accessibility in the human genome. *Cell* **184**, 5985–6001.e5919 (2021).
- Yoshimura, S. H. & Hirano, T. HEAT repeats - versatile arrays of amphiphilic helices working in crowded environments? *J. Cell Sci.* **129**, 3963–3970 (2016).
- Ghosh, S. G. et al. Biallelic hypomorphic mutations in HEATR5B, encoding HEAT repeat-containing protein 5B, in a neurological syndrome with pontocerebellar hypoplasia. *Eur. J. Hum. Genet.* **29**, 957–964 (2021).
- Wanker, E. E., Ast, A., Schindler, F., Trepte, P. & Schoeogl, S. The pathobiology of perturbed mutant huntingtin protein-protein interactions in Huntington's disease. *J. Neurochem.* **151**, 507–519 (2019).
- Zhou, H. et al. Analysis of long non-coding RNA expression profiles in neonatal rats with hypoxic-ischemic brain damage. *J. Neurochem.* **149**, 346–361 (2019).
- Wang, Y. et al. Hypermethylation of the enolase gene (ENO2) in autism. *Eur. J. Pediatr.* **173**, 1233–1244 (2014).
- Lake, J. et al. Coding and noncoding variation in LRRK2 and Parkinson's disease risk. *Mov. Disord.* **37**, 95–105 (2021).
- Grover, S. et al. Replication of a novel Parkinson's locus in a European ancestry population. *Mov. Disord.* **36**, 1689–1695 (2021).
- Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
- Zhou, X. et al. The Chinese Parkinson's Disease Registry (CPDR): study design and baseline patient characteristics. *Mov. Disord.* **37**, 1335–1345 (2022).
- Postuma, R. B. et al. MDS clinical diagnostic criteria for Parkinson's disease. *Mov. Disord.* **30**, 1591–1601 (2015).
- Hughes, A. J., Daniel, S. E., Kilford, L. & Lees, A. J. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J. Neurol. Neurosurg. Psychiatry* **55**, 181–184 (1992).
- Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
- Van der Auwera, G. A. et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinform.* **43**, 11 10 11–11 10 33 (2013).

36. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
37. Browning, B. L., Zhou, Y. & Browning, S. R. A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am. J. Hum. Genet.* **103**, 338–348 (2018).
38. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
39. Linderman, M. D. et al. Analytical validation of whole exome and whole genome sequencing for clinical applications. *BMC Med. Genomics* **7**, 20 (2014).
40. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genome-wide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
41. Higgins, J. P., Thompson, S. G., Deeks, J. J. & Altman, D. G. Measuring inconsistency in meta-analyses. *BMJ* **327**, 557–560 (2003).
42. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol.* **11**, e1004219 (2015).
43. Genomes Project, C. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
44. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
45. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
46. Brown, B. C., Asian Genetic Epidemiology Network Type 2 Diabetes, C., Ye, C. J., Price, A. L. & Zaitlen, N. Transethnic genetic-correlation estimates from summary statistics. *Am. J. Hum. Genet.* **99**, 76–88 (2016).
47. Buniello, A. et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).
48. Qi, T. et al. Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat. Commun.* **9**, 2282 (2018).
49. Yavorska, O. O. & Burgess, S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int. J. Epidemiol.* **46**, 1734–1739 (2017).
50. Choi, S. W. & O'Reilly, P. F. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience* **8**, giz082 (2019).
51. Wray, N. R., Yang, J., Goddard, M. E. & Visscher, P. M. The genetic interpretation of area under the ROC curve in genomic profiling. *PLoS Genet.* **6**, e1000864 (2010).

## ACKNOWLEDGEMENTS

The authors are grateful to the individuals with PD who participated in this study. We would thank all members of the Parkinson's Disease & Movement Disorders Multicenter Database and Collaborative Network in China. We also thank the research participants and employees of participating centres and study groups for making this work possible. This study was supported by the National Key Research and Development Program of China (Grant No. 2016YFC1306000), the Hunan Innovative Province Construction Project (Grant No. 2019SK2335), the Science and Technology Major Project of Hunan Provincial Science and Technology Department (Grant No. 2021SK1010), the National Natural Science Foundation of China (Grant No. 81430023), and the Key Research and Development Project of Ministry of Science and Technology of China (2016YFC0900802). This work was supported in part by the Intramural Research Programs of the National Institute on Aging (NIA) part of the National Institutes of Health, Department of Health and Human Services (1ZIA-NS003154, Z01-AG000949-02, Z01-AG000535, and Z01-ES101986).

<sup>1</sup>Department of Neurology, Xiangya Hospital, Central South University, 410008 Changsha, Hunan, China. <sup>2</sup>National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, 410008 Changsha, Hunan, China. <sup>3</sup>Department of Neurology, National Clinical Research Center for Geriatric Diseases, Xuanwu Hospital of Capital Medical University, 100053 Beijing, China. <sup>4</sup>Department of Neurology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, 200025 Shanghai, China. <sup>5</sup>Hunan Key Laboratory of Molecular Precision Medicine, Department of Oncology, Xiangya Hospital, Central South University, 410008 Changsha, Hunan, China. <sup>6</sup>Department of Geriatrics, Xiangya Hospital, Central South University, 410008 Changsha, Hunan, China. <sup>7</sup>Department of Neurology, The Second Xiangya Hospital, Central South University, 410011 Changsha, Hunan, China. <sup>8</sup>Department of Neurology, The Third Xiangya Hospital, Central South University, 410013 Changsha, Hunan, China. <sup>9</sup>Department of Neurology, Affiliated Brain Hospital of Nanjing Medical University, 210029 Nanjing, Jiangsu, China. <sup>10</sup>Department of Neurology, The First Affiliated Hospital of Zhengzhou University, 450047 Zhengzhou, Henan, China. <sup>11</sup>Department of Neurology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, 430022 Wuhan, Hubei, China. <sup>12</sup>Department of Neurology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, 430030 Wuhan, Hubei, China. <sup>13</sup>Department of Neurology, Renmin Hospital of Wuhan University, 430060 Wuhan, Hubei, China. <sup>14</sup>Department of Neurology, First Affiliated Hospital, Sun Yat-sen University, 510080 Guangzhou, Guangdong, China. <sup>15</sup>Department of Neurology, Zhujiang Hospital of Southern Medical University, 510280 Guangzhou, Guangdong, China. <sup>16</sup>Health Management Center, Hunan Provincial Brain Hospital, 410021 Changsha, Hunan, China. <sup>17</sup>Department of Neurology, Hunan Provincial Brain Hospital, 410021 Changsha, Hunan, China. <sup>18</sup>Department of Health Management Center, Xiangya Hospital, Central South University, 410008 Changsha, Hunan, China. <sup>19</sup>Institute of Aging & Tissue Regeneration, Renji Hospital, Shanghai Jiao Tong University School of Medicine, 200025 Shanghai, China. <sup>20</sup>Department of Neurology and Clinical Research Center of Neurological Disease, The Second Affiliated Hospital of Soochow University, 215004 Suzhou, Jiangsu, China. <sup>21</sup>Department of Neurology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, 510080 Guangzhou, Guangdong, China. <sup>22</sup>Department of Neurology, National Center of Gerontology, Beijing Hospital, 100005 Beijing, China. <sup>23</sup>Key Laboratory of Hunan Province in Neurodegenerative Disorders, Central South University, 410008 Changsha, Hunan, China. <sup>24</sup>Department of Epidemiology and Health Statistics, Xiangya School of Public Health, Central South University, 410028 Changsha, Hunan, China.

## AUTHOR CONTRIBUTIONS

Dr Tang, Dr Guo and Dr Li had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Concept and design were done by B.T., J.Y., J.G., J.L., H.Z., H.P., and Z.L. Acquisition, analysis, and interpretation of data were done by H.P., Z.L., J.M., Y.L., Y.Z., X.Z., Y.X., Y.W., X.Z., R.H., Y.X., Q.Z., K.Y., Q.X., Q.S., J.W., X.Y., H.Z., C.W., L.L., W.L., X.W., X.D., T.W., Z.X., Z.Z., L.C., Q.W., Y.L., J.T., X.Z., S.P., C.W., J.D., C.L., L.W., H.C., L.S., H.J., X.W., H.T., D.L., S.X., X.C., J.L., P.C., J.T., Z.H., C.C., K.X., Z.Z., J.F., C.B., M.N., A.S., J.L., P.C., H.Z., J.L., J.G., J.Y., and B.T. H.P., Z.L., J.Y., and B.T. drafted the original manuscript. All authors were involved in reviewing and editing the manuscript, and approved the manuscript for submission. Funding was obtained by B.T., J.G., and X.C. H.P. and Z.L. contributed equally to this work.

## COMPETING INTERESTS

Mike A. Nalls's declare no competing non-financial interests but the following competing financial interests: Mike A. Nalls's participation in this project was part of a competitive contract awarded to Data Tecnica International LLC by the National Institutes of Health to support open science research, he also currently serves on the scientific advisory board for Clover Therapeutics and is an advisor to Neuron23 Inc. Other authors report no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41531-023-00456-6>.

**Correspondence** and requests for materials should be addressed to Jian Yang or Beisha Tang.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

<sup>25</sup>Department of Social Medicine and Health Management, Xiangya School of Public Health, Central South University, 410028 Changsha, Hunan, China. <sup>26</sup>The Department of Dermatology, Xiangya Hospital, Central South University, 410008 Changsha, Hunan, China. <sup>27</sup>Centre for Medical Genetics & Hunan Key Laboratory of Medical Genetics, School of Life Sciences, Central South University, 410012 Changsha, Hunan, China. <sup>28</sup>Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore 308232, Singapore. <sup>29</sup>Molecular Genetics Section, Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892, USA. <sup>30</sup>Center for Alzheimer's and Related Dementias, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892, USA. <sup>31</sup>Data Tecnica International, Washington, DC 20037, USA. <sup>32</sup>Diseases & Population (DaP) Geninfo Lab, School of Life Sciences, Westlake University, 310024 Hangzhou, Zhejiang, China. <sup>33</sup>School of Life Sciences, Westlake University, 310024 Hangzhou, Zhejiang, China. <sup>35</sup>These authors contributed equally: Hongxu Pan, Zhenhua Liu. \*A list of authors and their affiliations appears at the end of the paper. ✉email: jian.yang@westlake.edu.cn; bstang7398@163.com

## THE PARKINSON'S DISEASE & MOVEMENT DISORDERS MULTICENTER DATABASE AND COLLABORATIVE NETWORK IN CHINA (PD-MDCNC)

Zhenhua Liu<sup>1,2,35</sup>, Jinghong Ma<sup>3</sup>, Yuanyuan Li<sup>4</sup>, Yuwen Zhao<sup>1</sup>, Qian Xu<sup>1</sup>, Qiyang Sun<sup>6</sup>, Junling Wang<sup>1</sup>, Xinxiang Yan<sup>1</sup>, Hainan Zhang<sup>7</sup>, Chunyu Wang<sup>7</sup>, Lifang Lei<sup>8</sup>, Weiguo Liu<sup>9</sup>, Xuejing Wang<sup>10</sup>, Xuebing Ding<sup>10</sup>, Tao Wang<sup>11</sup>, Zheng Xue<sup>12</sup>, Zhentao Zhang<sup>13</sup>, Ling Chen<sup>14</sup>, Qing Wang<sup>15</sup>, Yonghong Liu<sup>16</sup>, Jiayu Tang<sup>17</sup>, Chaodong Wang<sup>3</sup>, Chunfeng Liu<sup>20</sup>, Lijuan Wang<sup>21</sup>, Haibo Chen<sup>22</sup>, Lu Shen<sup>1</sup>, Hong Jiang<sup>1</sup>, Jun Liu<sup>4</sup>, Piu Chan<sup>3</sup>, Jinchen Li<sup>34</sup>, Jifeng Guo<sup>1</sup> and Beisha Tang<sup>1</sup>

<sup>34</sup>National Clinical Research Centre for Geriatric Disorders, Department of Geriatrics, Xiangya Hospital, Central South University, Changsha, China.