nature aging

Letter

https://doi.org/10.1038/s43587-024-00760-7

Exome sequencing in Asian populations identifies low-frequency and rare coding variation influencing Parkinson's disease risk

Received: 1 May 2024

Accepted: 24 October 2024

Published online: 21 November 2024

Check for updates

A list of authors and their affiliations appears at the end of the paper

Parkinson's disease (PD) is an incurable, progressive and common movement disorder that is increasing in incidence globally because of population aging. We hypothesized that the landscape of rare, protein-altering variants could provide further insights into disease pathogenesis. Here we performed whole-exome sequencing followed by gene-based tests on 4,298 PD cases and 5,512 controls of Asian ancestry. We showed that *GBA1* and *SMPD1* were significantly associated with PD risk, with replication in a further 5,585 PD cases and 5,642 controls. We further refined variant classification using in vitro assays and showed that *SMPD1* variants with reduced enzymatic activity display the strongest association (<44% activity, odds ratio (OR) = 2.24, $P = 1.25 \times 10^{-15}$) with PD risk. Moreover, 80.5% of *SMPD1* carriers harbored the Asian-specific p.Pro332Arg variant (OR = 2.16; $P = 4.47 \times 10^{-8}$). Our findings highlight the utility of performing exome sequencing in diverse ancestry groups to identify rare protein-altering variants in genes previously unassociated with disease.

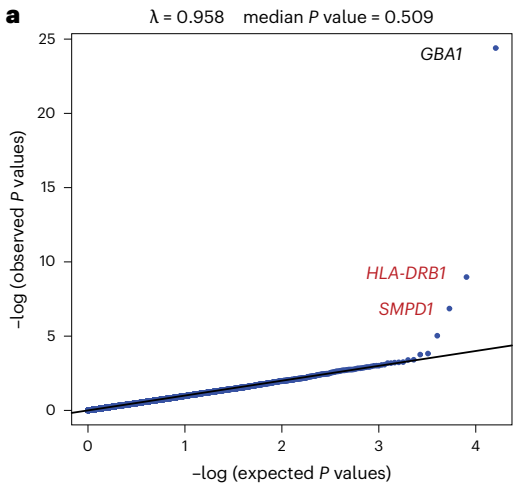
Parkinson's disease (PD) is a common neurodegenerative disease that is unrelenting and progressive. It presents as a movement disorder characterized by slowness of movement, tremors, stiffness and difficulty with balance, together with a broad array of non-motor symptoms[1,2]. Although levodopa and other adjunctive medications alleviate the symptoms of PD, no cure has yet been found. The incidence of PD is expected to rise rapidly in the coming decades because of the aging population. There is, thus, a pressing need for new therapeutic approaches, with the hope that biological insights arising from molecular genetic studies could illuminate these otherwise elusive drug targets.

Owing to reports of positive family histories in many affected individuals, PD has long been suspected to be heritable. Genome-wide association studies (GWASs) of PD have identified close to 100 loci associated with disease risk[3–9]. Although these studies highlighted key genes and biological pathways that play a role in the pathogenesis of PD, they also underscore the very heterogeneous nature of the disease in terms of causative mechanisms.

The genomic/risk loci identified by GWASs of PD mostly have modest effect sizes (odds ratios (ORs) < 1.4) and were often located in non-coding regions of the human genome[9]. Their non-coding nature makes pinpointing the effector genes challenging. In contrast, analysis of protein-altering genetic variants using whole-exome sequencing (WES) has identified causative genes and facilitated advances in the treatment of other diseases through the identification and development of new drug targets[10,11]. In studies on familial, early-onset PD, several genes containing rare functionally deficient genetic variants have been found[12–26]. Such variants were highly penetrant and were seldom seen in patients with sporadic PD[27]. In the case of adult-onset sporadic PD, the effect of individual rare genetic variants remains difficult to assess, but the aggregation of rare variant counts within each gene could allow for detection of statistically significant differences in mutational 'burden' between cases and controls. These gene-based tests of mutational burden are often limited by constraints in accuracy of pathogenicity prediction using bioinformatic algorithms[28] because not all variants that are predicted to impair protein function actually

e-mail: khorcc@gis.a-star.edu.sg; gnrtek@sgh.com.sg; zhenxun.wang@duke-nus.edu.sg; jianee.foo@ntu.edu.sg

https://doi.org/10.1038/s43587-024-00760-7



b

Population	Case Carrier	Case Non-carrier	Control Carrier	Control Non-carrier	No. of variants	OR (95% CI)	*P* value
SG/MAL	85	2,526	53	3,691	22	2.343 (1.637–3.38)	
HK	3	67	12	574	5	2.139 (0.378–8.203)	
KR	12	1,405	3	1,037	8	2.951 (0.794–16.337)	
TW	0	200	1	141	1	0.24 (0.01–5.82)	
CMH_discovery						2.32 (1.682–3.201)	1.37×10^{-7}
Zhao	114	3,765	61	2,870	41	1.425 (1.031–1.985)	
Robak	14	1,153	16	1,669	15	1.266 (0.57–2.781)	
Alcalay-Boston	11	539	10	1,022	10	2.085 (0.798–5.514)	
CMH_replication						1.453 (1.105–1.91)	7.05×10^{-3}
CMH_all						1.766 (1.433–2.175)	4.94×10^{-8}

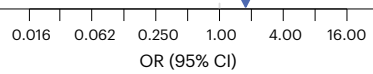


Fig. 1 | Exome-wide significant association of *GBA1*, *HLA-DRB1* and *SMPD1* in the Discovery exome sequencing study. a, Quantile–Quantile plot of observed and expected CMH *P* for autosomal genes with OR > 1. **b**, Forest plot of *SMPD1* association in assessed populations. CMH without continuity correction was used across strata. Data are presented as OR and 95% CI. All *P* values presented are unadjusted. Exome-wide significance was set at $P < 2.5 \times 10^{-6}$ (two-tailed), taking into account multiple hypothesis testing correction for an estimated 20,000 protein-coding genes in the human genome.

do so. Conversely, variants that are predicted to be benign could result in a loss of function of the encoded protein.

The objective of this study was to assess the contribution of rare coding sequence genetic variants to the pathogenesis of PD. We performed WES on 9,810 participants of Asian ancestry (4,298 PD cases and 5,512 unaffected controls) from Singapore (SG), Malaysia (MAL), Hong Kong (HK), South Korea (KR) and Taiwan (TW). We attempted to replicate two significantly associated genes in a further 11,227 participants (5,585 PD cases and 5,642 controls) from three independently ascertained collections of Asian and European ancestry. We used functional assays to validate and refine gene-based mutational burden tests.

Results

Rare coding variants identified by WES

We recruited 4,915 PD cases together with 5,971 ancestry-matched and geographically matched controls from five regions across East Asia (SG, MAL, HK, KR and TW) and analyzed a total of 4,298 PD cases and 5,512 controls by WES after stringent quality control filtering (total $n = 9{,}810$; Supplementary Tables 1–3 and Supplementary Fig. 1). Only samples confirmed by principal component analysis (PCA) to be of Han Chinese and South Korean ancestries were retained (Supplementary Figs. 2 and 3). We identified a total of 721,643 non-synonymous variants across 17,396 autosomal genes, of which 701,048 were rare (minor allele frequency (MAF) < 1%) in the sequenced dataset as well as in publicly available datasets (Genome Aggregation Database (gnomAD) East Asian samples). After annotation with the PolyPhen-2 and LOFTEE algorithms, 367,878 potentially pathogenic variants (hereafter termed 'qualifying variants') remained for further analysis.

Gene-level association testing revealed exome-wide significant PD associations with carriage of qualifying variants in *GBA1*, *HLA-DRB1* and *SMPD1* (Fig. 1, Extended Data Fig. 1 and Supplementary Table 4). All three genes were expressed in the human midbrain (Supplementary Results, Supplementary Tables 5–7 and Supplementary Figs. 4–6).

GBA1 and *SMPD1* qualifying variants are associated with higher PD risk

At *GBA1*, 128 of 4,298 PD cases (2.98%) carried qualifying variants compared to 16 of 5,512 controls (0.29%; OR = 10.28, 95% confidence interval (CI) = 5.96–17.75, $P = 4.07 \times 10^{-25}$; Supplementary Table 8 and Extended

Table 1 | *SMPD1* coding variants from the Discovery dataset

	Position (Hg37)	Amino acid substitution	gnomAD 2.1.1 East Asian frequency		Case		Control		Pathogenicity prediction	Functional score
			Exomes	Genomes	No. of heterozygotes	No. of homozygotes	No. of heterozygotes	No. of homozygotes		
Rare deleterious variants	chr11.6411889.C.T	p.Gln21Ter	-	-	0	0	1	0	Stop-gain	25.00% (24.90–25.10%)
	chr11.6412063.C.T	p.Arg79Trp	-	-	1	0	0	0	Probably damaging	69.60% (69.30–69.80%)
	chr11.6413101.C.T	p.Ala269Val	4.00×10^{-4}	6.00×10^{-4}	1	0	2	0	Probably damaging	98.80% (92.30–105.70%)
	chr11.6413106.C.A	p.Pro271Thr	-	-	0	0	1	0	Damaging	76.40% (75.30–77.50%)
	chr11.6413134.A.ACATCCCCG	p.His284SerfsTer18	-	-	2	0	0	0	Frameshift insertion	5.60% (5.40–5.90%)
	chr11.6413137.TC.T	p.Ala283HisfsTer16	-	-	0	0	1	0	Frameshift deletion	9.30% (9.20–9.40%)
	chr11.6413139.C.A	p.Pro282Thr	2.00×10^{-4}	-	2	0	0	0	Damaging	9.70% (9.70–9.70%)
	chr11.6413142.G.A	p.Ala283Thr	1.00×10^{-4}	-	1	0	0	0	Damaging	11.10% (11.00–11.20%)
	chr11.6413181.C.T	p.Arg296Trp	5.44×10^{-5}	-	3	0	0	0	Damaging	87.10% (85.50–88.60%)
	chr11.6413187.C.G	p.Leu298Val	-	-	1	0	0	0	Damaging	100.90% (99.90–101.80%)
	chr11.6413235.G.A	p.Val314Met	0	0	1	0	0	0	Damaging	55.30% (54.80–55.80%)
	chr11.6413278.G.C	p.Ser328Thr	-	-	1	0	0	0	Probably damaging	34.20% (34.00–34.30%)
	chr11.6413290.C.G	p.Pro332Arg	5.60×10^{-3}	1.30×10^{-3}	69	1	48	0	Damaging	30.40% (30.30–30.50%)
	chr11.6413296.T.C	p.Ile334Thr	0	0	0	0	1	0	Damaging	78.80% (78.40–79.10%)
	chr11.6414486.C.T	p.Arg378Cys	0	-	1	0	0	0	Damaging	37.10% (36.70–37.60%)
	chr11.6414487.G.A	p.Arg378His	2.00×10^{-4}	0	2	0	3	0	Damaging	63.80% (63.60–64.00%)
	chr11.6414492.A.T	p.Ile380Phe	-	-	1	0	0	0	Damaging	41.00% (40.40–41.50%)
	chr11.6414543.A.G	p.Asn397Asp	-	-	0	0	1	0	Probably damaging	67.80% (67.30–68.30%)
	chr11.6414558.G.A	p.Ala402Thr	0	-	1	0	0	0	Damaging	64.20% (64.00–64.40%)
	chr11.6414877.C.T	p.His432Tyr	5.44×10^{-5}	-	1	0	0	0	Probably damaging	111.60% (111.40–111.90%)
	chr11.6414911.G.A	p.Arg443Gln	4.00×10^{-4}	6.00×10^{-4}	5	0	4	0	Damaging	24.00% (24.00–24.00%)
	chr11.6415556.T.C	p.Tyr539His	-	-	0	0	1	0	Damaging	10.20% (10.20–10.20%)
	chr11.6415622.C.T	p.Arg561Cys	0	-	0	0	1	0	Damaging	47.50% (47.30–47.80%)
	chr11.6415712.C.T	p.Arg591Cys	0	0	3	0	1	0	Damaging	40.60% (40.50–40.60%)
	chr11.6415730.G.A	p.Ala597Thr	-	-	2	0	0	0	Probably damaging	50.80% (50.70–51.00%)
	chr11.6415736.C.T	p.Leu599Phe	-	-	0	0	1	0	Damaging	27.90% (27.80–28.10%)
	chr11.6415746.G.A	p.Arg602His	0	-	1	0	2	0	Damaging	29.50% (29.20–29.80%)
	chr11.6415769.C.T	p.Arg610Cys	0	0	0	0	1	0	Damaging	73.70% (73.50–73.80%)

https://doi.org/10.1038/s43587-024-00760-7

https://doi.org/10.1038/s43587-024-00760-7

Table 1 (continued) | *SMPD1* coding variants from the Discovery dataset

	Position (Hg37)	Amino acid substitution	gnomAD 2.1.1 East Asian frequency		Case		Control		Pathogenicity prediction	Functional score
			Exomes	Genomes	No. of heterozygotes	No. of homozygotes	No. of heterozygotes	No. of homozygotes		
Rare benign variants	chr11.6411878.G.A	p.Arg17Gln	1.00×10^{-4}	-	3	0	2	0	NA	119.60% (118.70–120.60%)
	chr11.6411917.T.G	p.Leu30Arg	-	-	1	0	1	0	NA	77.10% (76.80–77.50%)
	chr11.6411958.G.C	p.Ala44Pro	2.00×10^{-4}	6.00×10^{-4}	0	0	1	0	Benign	126.80% (126.30–127.30%)
	chr11.6412022.C.T	p.Ser65Phe	-	-	0	0	1	0	Benign	99.60% (98.00–101.30%)
	chr11.6412054.A.C	p.Ile76Leu	-	-	1	0	0	0	Benign	102.20% (101.40–102.90%)
	chr11.6413089.G.C	p.Gly265Ala	-	-	2	0	2	0	Benign	57.10% (57.00–57.30%)
	chr11.6413277.A.G	p.Ser328Gly	-	-	1	0	0	0	Benign	82.30% (81.70–82.90%)
	chr11.6413305.A.G	p.Asn337Ser	5.44×10^{-5}	-	2	0	0	0	Benign	51.80% (51.50–52.00%)
	chr11.6413377.G.A	p.Arg361His	5.44×10^{-5}	0	2	0	1	0	Benign	68.30% (68.00–68.50%)
	chr11.6415259.G.A	p.Gly492Ser	2.00×10^{-4}	6.00×10^{-4}	0	0	3	0	Benign	49.10% (48.60–49.60%)
	chr11.6415583.C.T	p.Pro548Ser	5.44×10^{-5}	-	0	0	1	0	Benign	64.80% (64.50–65.20%)
	chr11.6415590.C.T	p.Thr550Ile	-	-	1	0	0	0	Benign	60.20% (59.60–60.90%)
	chr11.6415634.G.A	p.Asp565Asn	1.00×10^{-4}	-	1	0	1	0	Benign	105.30% (104.90–105.70%)
	chr11.6415639.G.T	p.Met566Ile	2.00×10^{-4}	-	1	0	3	0	Benign	70.40% (70.30–70.60%)
	chr11.6415704.C.T	p.Thr588Met	0.001	0.0013	9	0	10	0	Benign	62.30% (59.30–65.50%)
Common variants	chr11.6415463.G.A	p.Gly508Arg	0.137	0.153	1,018	83	1,329	115	Damaging	55.60% (55.40–55.80%)
	chr11.6415539.C.T	p.Pro533Leu	0.007	0.014	76	1	76	0	Damaging	72.10% (71.90–72.30%)

https://doi.org/10.1038/s43587-024-00760-7

Data Fig. 2). The carriage of 22 East Asian–specific rare deleterious *GBA1* variants was strongly and consistently associated with increased risk of PD (OR = 25.83, $P = 6.19 \times 10^{-11}$) compared to non-carriers (Supplementary Table 9).

Turning to *SMPD1*, 100 of 4,298 PD cases (2.33%) carried qualifying variants compared to 69 of 5,512 controls (1.25%; OR = 2.32, 95% CI = 1.68–3.20, $P = 1.37 \times 10^{-7}$; Table 1, Fig. 1 and Supplementary Tables 10 and 11). The *SMPD1* association was driven by one particular variant (p.Pro332Arg) in the Discovery cohort (allele frequency: PD cases = 0.8%, controls = 0.4%; OR = 2.39; $P = 5.01 \times 10^{-6}$). When we excluded p.Pro332Arg from the analysis, the significant association between burden of *SMPD1* qualifying variants and increased PD risk remained (OR = 2.31, $P = 0.00369$), suggesting that the association between rare, independent genetic variants in *SMPD1* was not confined to p.Pro332Arg alone.

We did not observe significant differences in effect sizes between males and females at *GBA1* (males: OR = 10.03, 95% CI = 4.53–22.21; females: OR = 9.90, 95% CI = 4.75–20.62; Phet = 0.981) or *SMPD1* (males: OR = 2.63, 95% CI = 1.68–4.12, females: OR = 1.96, 95% CI = 1.22–3.14; Phet = 0.376).

We did not observe significant association at genes previously implicated in autosomal dominant, recessive[29–31] or X-linked PD[32] (Supplementary Tables 12 and 13) or genes identified by a recent WES study in individuals of European ancestry[27] (Supplementary Table 14). We also did not observe significant PD associations ($P < 0.00058$, Bonferroni correction for 86 tested genes) in genes involved in lysosomal storage disorders (Supplementary Table 15)[33,34].

Validation of qualifying *SMPD1* variants associated with increased PD risk

As the association between rare *GBA1* variants and increased PD risk has been confirmed[26], we focused on evaluating whether the association between rare *SMPD1* variants and PD could be replicated in additional independent samples. We pursued validation in 5,585 PD cases and 5,642 controls from three studies of Asian and European ancestry (Supplementary Table 16). Participants from these three studies have undergone whole-exome or whole-genome sequencing, with documented equal re-sequencing of cases and controls[35], and had *SMPD1* data available. Stratified meta-analysis of the three Replication collections showed that carriers of *SMPD1* qualifying variants remain overrepresented in PD cases (139 carriers in 5,585 PD cases, 2.5%) compared to controls (87 carriers in 5,642 controls, 1.5%) (OR = 1.45, $P = 0.0071$). When all Discovery and Replication samples were meta-analyzed (9,883 PD cases and 11,154 controls across seven independent strata), genome-wide significant association between carriage of *SMPD1* qualifying variants and susceptibility to PD was observed (OR = 1.77, $P = 4.94 \times 10^{-8}$; Fig. 1 and Supplementary Table 17).

We observed significant heterogeneity (I^2 index for heterogeneity = 79.6%, $P = 0.027$) in OR between the Discovery (OR = 2.32) and Replication (OR = 1.45) cohorts, which could be due to misclassification of qualifying variants by the PolyPhen-2/LOFTEE algorithm. Resolving this would necessitate the design of functional assays to measure the activity of each *SMPD1* allele for comparison against wild-type *SMPD1*. Alleles that demonstrate significant loss of function

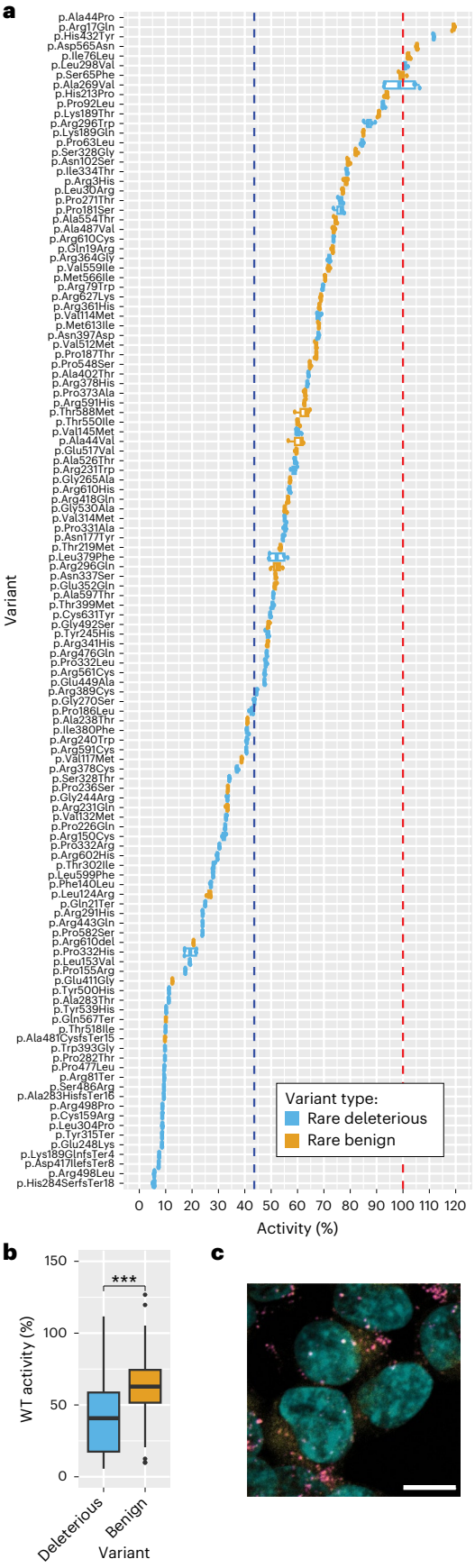


Fig. 2 | ASM activity levels of *SMPD1* variants in Discovery and Replication datasets. a, Activity levels are presented as percentage of wild-type ASM activity; error bars represent standard deviation. Each variant was assayed in technical replicates of 3–6. **b**, Predicted deleterious variants (75 variants) have significantly lower ASM activity than predicted benign variants (47 variants). Median, 25th–75th quartiles; whiskers of 1.5× IQR shown in boxplot. *** Two-sided *t*-test $P < 0.001$. Red dotted line: 100% activity of wild-type variants. Blue dotted line: activity level (43.58%) significantly associated with variant carrier frequency in case and control. Details are provided in Supplementary Table 17. **c**, SMPD1 wild-type protein localizes to the lysosome. Representative image from three captured fields of view is shown (Supplementary Fig. 11). Yellow signal: SMPD1. Magenta signal: lysosome. Cyan signal: nuclei. Scale bar, 10 µm. IQR, interquartile range; WT wild-type.

compared to wild-type would be included in the revised gene-based burden test. Conversely, alleles that show indistinguishable functional activity from wild-type *SMPD1* would be excluded from gene-based burden tests regardless of PolyPhen-2/LOFTEE prediction. *SMPD1* encodes for acid sphingomyelinase (ASM), an enzyme responsible for breaking down sphingomyelin to ceramide and phosphorylcholine[36]. Complete deficiency of ASM leads to Niemann–Pick disease type A/B (NPD-A/B). We designed enzymatic assays to measure ASM activity for 125 *SMPD1* genetic variants detected across Discovery and Replication samples (Fig. 2a,b, Supplementary Figs. 7–10, Supplementary Table 17 and Supplementary Results). For the assays, we created 126 HEK293T stable cell lines expressing each *SMPD1* genetic variant and 'wild-type' ASM. Confocal microscopy confirmed correct localization of wild-type ASM to the lysosomes in the HEK293T cell line (Fig. 2c and Supplementary Fig. 11).

The functional spectrum of ASM activity for all tested *SMPD1* variants ranged from normal (>90% of wild-type activity) to severe loss of function (<20% of wild-type activity). Two of the variants tested are causative of NPD-A/B (*SMPD1* p.Leu304Pro (or p.Leu302Pro[37]) and p.Arg498Leu[38]) and, thus, served as positive controls for our experiment. Unsurprisingly, both variants ranked in the bottom 10th percentile of ASM activity (p.Leu304Pro: 8.7% activity, p.Arg498Leu: 5.6% activity) in our assay. Across 125 variants tested, we observed that PolyPhen-2/LOFTEE-predicted damaging variants correlated significantly, albeit imperfectly, with lower ASM activity levels compared to predicted benign variants (*t*-test $P = 1.11 \times 10^{-5}$). We observed that carriage of *SMPD1* variants with less than 44% enzymatic activity (corresponding to >56% loss of function) showed the strongest association with PD risk in both the Discovery (OR = 2.37, $P = 4.35 \times 10^{-7}$) and Replication (OR = 2.18, $P = 4.80 \times 10^{-10}$) cohorts and the combined Discovery and Replication cohorts (OR = 2.24, $P = 1.25 \times 10^{-15}$; Fig. 3a, Extended Data Figs. 3 and 4, Supplementary Fig. 12 and Supplementary Results). In our East Asian Discovery cohort, the p.Pro332Arg variant (30.4% activity) was the most prevalent, observed in 70 of 87 (80.5%) *SMPD1* case carriers with less than 44% enzymatic activity relative to wild-type (Fig. 3b). p.Pro332Arg alone showed genome-wide significant association with PD risk (OR = 2.16; $P = 4.47 \times 10^{-8}$) in the Discovery and Replication cohorts combined (Supplementary Results).

Forty-four of the 125 *SMPD1* variants assayed were reported in patients with NPD-A or NPD-B and were also classified in ClinVar as pathogenic ($n = 18$), benign ($n = 5$) or uncertain significance ($n = 21$) (Supplementary Table 17). Fourteen of 18 (77.78%) ClinVar pathogenic variants reported in patients with NPD-A/B had less than 44% enzymatic activity relative to wild-type SMPD1, thus confirming that the functionally deficient set is enriched for pathogenic variants.

PD onset age and polygenic risk scores of *SMPD1* and *GBA1* carriers

We sought to determine if carriers of *SMPD1* or *GBA1* variants from our Discovery samples had an earlier age of PD diagnosis compared to non-carriers. Carriers of *GBA1* qualifying variants were diagnosed with PD almost 7 years younger than non-carriers, but the effect was smaller for carriers of *SMPD1* qualifying variants and carriers of functionally deficient *SMPD1* variants (activity levels ≤43.58%; carriers: median = 60 years of age, range = 32–86 years of age; non-carriers: median = 62 years of age, range = 22–92 years of age; *t*-test $P = 0.216$) (Extended Data Fig. 5). We examined the polygenic risk score profile of 87 *SMPD1* functionally deficient variant carriers calculated using 67 GWAS-identified single-nucleotide polymorphisms (SNPs) selected in our East Asian GWAS[39] and observed no significant difference between PD cases and controls, suggesting that polygenic risk score was not a modifier of risk among *SMPD1* functionally deficient variant carriers (Extended Data Fig. 6).

Rare coding variants in genes associated with decreased PD risk

We observed exome-wide significant association of qualifying variants in *HLA-DRB1* with decreased PD risk. Qualifying variants in *HLA-DRB1* were observed in 15 of 4,298 PD cases (0.35%) compared to 73 of 5,512 controls (1.32%; OR = 0.20, 95% CI = 0.11–0.35, $P = 1.05 \times 10^{-9}$; Supplementary Table 18 and Extended Data Fig. 7). This association was largely driven by lead variant p.Ala14Pro (OR = 0.13, 95% CI = 0.07–0.26, $P = 4.71 \times 10^{-11}$) located in the signal peptide of the full length HLA-DRB1 protein and not linked to any classical *HLA-DRB1* allele. This variant was present in 10 PD cases (0.23%) and 68 controls (1.23%) in our Discovery cohort and appeared to be independent of GWAS-associated SNPs[40,41] at the *HLA-DRB1* locus, including rs601945, rs504594 or rs3104413, and amino acid 13 ($r^2 = 0$). However, given the difficulty in calling and validating variants in the HLA region by short-read sequencing, further work will be needed to ascertain the association of altered epitope binding or antigen presentation effects of these *HLA-DRB1* variants with reduced PD risk[40,41]. We also observed a nominally significant association at *ATP5B* (OR = 0.19, 95% CI = 0.08–0.43, $P = 9.46 \times 10^{-6}$), but this association was not consistent in the South Korean dataset (Extended Data Fig. 7).

Monogenic gene mutations are rare in East Asian sporadic PD cases

We investigated how many East Asian sporadic PD cases carried known or likely pathogenic variants (ClinVar) in monogenic PD genes. In autosomal recessive PD genes, only one PD case (onset age = 44 years) was homozygous for pathogenic variant *PRKN* p.Cys441Arg, which was reported for autosomal recessive juvenile PD (ClinVar accession: VCV000963190). Sixteen PD cases and six controls (OR = 3.845, $P = 0.00375$) were homozygotes or compound heterozygotes for qualifying variants excluding those listed as non-pathogenic in ClinVar (benign, conflicting interpretations or uncertain significance) in *PRKN* (two cases), *ATP13A2* (one case), *DNAJC6* (one control) and *VPS13C* (13 cases and five controls) (Supplementary Table 12). These carriers had PD onset ranging from 28 years to 79 years of age (average = 59.5 ± 14.1 years, median = 60 years; $n = 14$; two missing).

Although there was no significant enrichment of qualifying variants in autosomal dominant PD genes, we identified carriers of pathogenic *LRRK2* variants p.Arg1441Cys (four cases, disease onset age = 56 ± 4.3 years) and p.Arg1067Gln (four cases and two controls) that activate Rab10 phosphorylation[42] (Supplementary Tables 12 and 13).

Seven of the 4,298 cases in our study carried known or potential pathogenic variants in known monogenic PD genes, and another 15 were potential homozygotes or compound heterozygotes for qualifying variants in monogenic PD genes. This is consistent with previous reports that these monogenic PD mutations account for a very small percentage (~0.5%) of sporadic PD cases.

Protein-truncating variants in East Asian PD cases and controls

Inaccurate functional prediction of missense variants may dilute the association signals in genes with highly penetrant loss-of-function variants with large effects. To assess this, we conducted gene-based association testing in the Discovery cohort on protein-truncating variants (PTVs) predicted to cause loss of function[43]. No gene reached exome-wide significance, although *GBA1* remained strongly associated (OR = 10.8, $P = 0.0024$; Supplementary Results and Supplementary Table 19). We observed PTVs in the recently reported *ITSN1* gene[44] in only three cases in our Discovery cohort ($P = 0.093$).

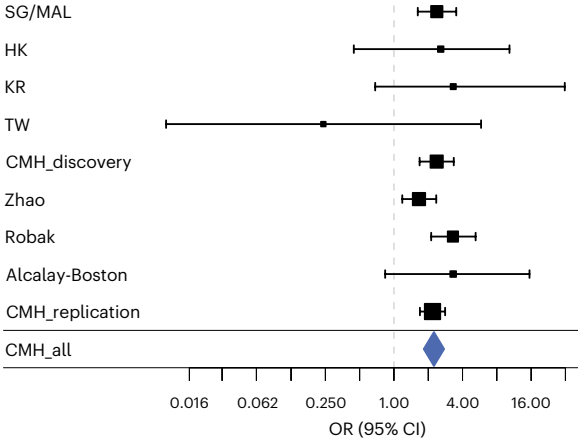
Discussion

Our observations from WES, replication in independent, cross-ancestry samples and functional enzymatic assays suggest that heterozygous carriage of deficient *SMPD1* alleles is associated with 2.2-fold higher

https://doi.org/10.1038/s43587-024-00760-7

a

Population	Case		Control		No. of variants	OR (95% CI)	*P* value
	Carrier	Non-carrier	Carrier	Non-carrier			
SG/MAL	75	2,536	46	3,698	10	2.377 (1.619–3.523)	
HK	3	67	10	576	3	2.574 (0.444–10.33)	
KR	9	1,408	2	1,038	4	3.316 (0.684–31.598)	
TW	0	200	1	141	1	0.24 (0.01–5.82)	
CMH_discovery						2.372 (1.679–3.351)	4.35×10^{-7}
Zhao	115	3,764	53	2,878	29	1.659 (1.183–2.351)	
Robak	70	1,097	32	1,653	8	3.295 (2.123–5.212)	
Alcalay-Boston	7	543	4	1,028	4	3.31 (0.837–15.486)	
CMH_replication						2.181 (1.691–2.813)	4.80×10^{-10}
CMH_all						2.244 (1.828–2.755)	1.25×10^{-15}



b

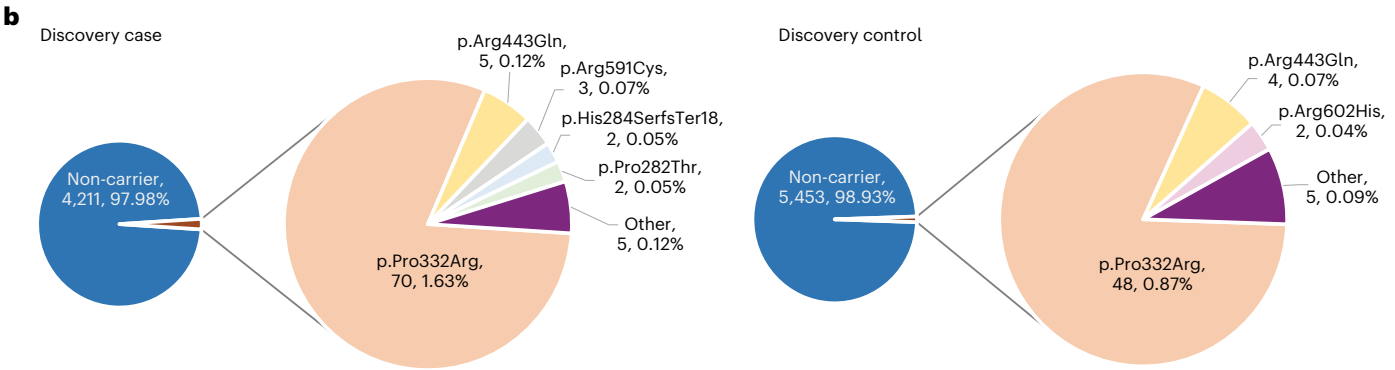


Fig. 3 | Replication of *SMPD1* association in Chinese and European datasets for rare variants with ≤43.58% ASM activity level. a, Forest plot of *SMPD1* in assessed populations. CMH without continuity correction was used across strata. Data are presented as OR and 95% CI. **b**, Distribution of rare variants with ≤43.58% ASM activity level in case and control from the Discovery cohort.

odds of PD. The top two genes emerging from WES, *GBA1* and *SMPD1*, are both involved in lysosomal storage diseases (Gaucher's disease[45] and NPD[46]) that present early in life with severe clinical symptoms.

GBA1 has been extensively evaluated since its identification as a PD risk gene[26], with patients carrying rare, protein-altering variants robustly associating with increased PD risk and more severe clinical phenotype[47]. Conversely, the association between protein-altering *SMPD1* variants and PD had been inconsistent. The first study linking *SMPD1* with PD reported that carriage of the p.Leu304Pro variant was associated with nine-fold increased odds of disease[48]. *SMPD1* p.Leu304Pro was a loss-of-function, founder mutation in individuals of Ashkenazi Jewish ancestry[48,49] and was not present in nearly all other ancestry groups, including East Asians[50]. Since then, other smaller studies have shown equivocal evidence in support for other *SMPD1* variants[33,34,48,51,52], and others have reported no significant association between *SMPD1* variants and susceptibility to PD[22,53–55]. *SMPD1* was not identified in the largest WES study of European ancestry samples, which reported strong association at *GBA1* and *LRRK2* (driven by the rare p.Gly2019Ser variant that is absent in Asian populations)[27]. We suspect that these inconsistent *SMPD1* reports were due to the limitations of the bioinformatic algorithms used to predict and identify 'damaging' genetic variants. Indeed, incorrect functional classification of genetic variants markedly reduces statistical power for rare variant burden tests[28]. The different *SMPD1* rare variants represented at different allele frequency spectra between Asian and European cohorts may also explain the heterogeneity observed in the combined Discovery and Replication analysis.

We used laboratory assays based on ASM activity to evaluate the functional effect of all 125 *SMPD1* alleles detected in this study. This assay was adequately controlled for, as it was able to correctly classify mutations known to cause NPD-A/B. The inclusion of only functionally impaired variants in the gene-based burden test is a major step forward from the current use of predictive bioinformatic algorithms to distinguish variants that impair protein function from 'benign' variants. Our analysis suggests that carrying one functionally defective copy of *SMPD1* was sufficient to associate with more than two-fold increased risk of sporadic, adult-onset PD. In contrast, complete loss of *SMPD1* function causes childhood-onset NPD-A/B. Bearing in mind that ASM encoded by *SMPD1* catalyzes the conversion of sphingomyelin to phosphocholine and ceramide[36], previous investigators observed significant differences in sphingomyelin and ceramide lipid species between postmortem brains from PD cases compared to brains from non-PD donors[56,57]. Taken together, the spectrum of functional *SMPD1* alleles detected in our large-scale sequencing study affirm a role for ASM in the pathobiology of neurodegenerative diseases.

Limitations

Our study has several limitations. First, we were unable to assess the effects of co-carrying functionally deficient variants in both *GBA1* and *SMPD1*, as we observed only eight individuals who carried protein-altering variants in both *GBA1* and *SMPD1*. Of these, two were PD cases who carried a known pathogenic variant in *GBA1* (p.Leu483Pro) with *SMPD1* (p.Pro332Arg). Their variable age of disease

onset (48 years and 62 years) is in keeping with the high genetic heterogeneity observed in sporadic PD. Future work involving more participants carrying mutations in both genes will be needed to resolve the question of possible interaction between both genes. Second, the concordance between our in vitro ASM activity assays and previously reported measurements based on dried blood spots was approximately 50%[58]. This discordance could reflect variability in laboratory techniques as well as opportunities to perform experiments in multiple biological replicates for more accurate ascertainment of enzymatic activity. Reassuringly, one such discordant variant (*SMPD1* p.Ala487Val) was re-assayed by a third party, and their conclusions agreed with our findings that this variant does not qualify as functionally deficient[59]. Finally, large sample sizes are typically required for rare variant association studies. In our current Discovery cohort, we had more than 90% power (Extended Data Fig. 8) to detect genes such as *GBA1* with large effect sizes (OR > 5) and at combined carrier frequencies of approximately 0.5% in the population and also genes such as *SMPD1* with smaller effect sizes (OR ~ 2) but at higher combined carrier frequencies of 1–1.5% in the population (contributed, in part, by the p.Pro332Arg variant). Larger multi-center studies will facilitate the discovery of genes with similar effect sizes to *GBA1* and *SMPD1* but present at lower carrier frequencies in the population.

Conclusion

From our study, we suggest that rare independent alleles in *SMPD1* that impair ASM activity are associated with increased risk of PD. Our findings will inform further work on understanding the role of sphingolipid and ceramide metabolism as well as the role of rare, functionally deficient alleles in associated genes in PD pathogenesis.

Methods

Study design and participants

Discovery exome sequencing. PD cases together with ancestry-matched and geographically matched controls were recruited by six independent study groups from five regions across East Asia (SG, MAL, HK, KR and TW). A total of 4,298 PD cases and 5,512 controls were included in the final analysis (Supplementary Tables 1 and 2). A diagnosis of PD was reached using the United Kingdom Parkinson's Society Brain Bank Criteria. The controls were healthy individuals who do not have any overt neurological diseases. Written informed consent was obtained from each participant in strict adherence to the Declaration of Helsinki. This study was approved by the ethics committees or institutional review boards of the respective institutions (SingHealth Centralized Institutional Review Board CIRB 2002/008/A and 2019/2334 and Nanyang Technological University Institutional Review Board IRB-2016-08-011). The study was conducted from 1 January 2016 to 31 December 2022. This study followed the Strengthening the Reporting of Genetic Association Studies (STREGA) guidelines.

Replication datasets. Published sequencing data from 5,585 PD cases and 5,642 controls from China, the United States and the International PD Genomics Consortium (IPDGC, samples from the United States and Europe) were evaluated for replication of significant findings from the Discovery WES analysis (Supplementary Table 16).

For the Chinese dataset, a total of 1,440 sporadic early-onset PD cases, 477 autosomal dominant or autosomal recessive PD cases, 1,962 sporadic late-onset PD cases and 2,931 controls from Changsha were included[33] (referred to as the 'Zhao' dataset/stratum). For the IPDGC dataset, a total of 691 sporadic early-onset PD cases, 465 early-onset PD cases with positive family history of PD and 1,679 controls of Northern and Western European ancestry were included[34] (referred to as the 'Robak' stratum). For the United States dataset, a total of 550 PD cases and 284 controls from New York were enrolled[58]. This dataset was enriched with an additional 748 controls from neighboring Boston[60] (referred to as the 'Alcalay-Boston' stratum). *SMPD1* variants analyzed in all 11,227 Replication samples were from sequencing data obtained using similar experimental methods from both PD cases and controls. No population databases (for example, gnomAD or UK Biobank) were used in lieu of sequenced controls, which circumvented the introduction of bias in coverage and variant calling.

Whole-exome capture and sequencing

Peripheral venous blood was drawn from each consenting participant, and DNA extraction was performed using routine laboratory methods. DNA for WES was sheared using a Covaris LE220 Focused-ultrasonicator, and libraries were prepared using NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs (NEB)). Exomes were captured using SeqCap EZ Human Exome Library version 3.0 kits or KAPA HyperExome Probes (both from Roche NimbleGen). Libraries from PD cases and controls were barcoded and sequenced together in batches of 96 samples using 2 × 151-bp paired-end reads on Illumina instruments (HiSeq 4000 and NovaSeq 6000 systems) to achieve an average of 63× coverage depth across captured target regions, with at least 95% of targeted exons covered by at least 10 reads. All samples were processed in the same laboratory.

Bioinformatics processing, variant calling and quality control filtering

For the Discovery WES dataset, raw DNA sequence reads from all samples (with case–control status blinded) were mapped to the human reference genome build hg19 (GRCh37) using the Burrows–Wheeler Aligner MEM algorithm (BWA version 0.7.17)[61] (Supplementary Fig. 1). Variant calling was performed using the Genome Analysis Tool Kit (GATK, version 3.7)[62] and the Picard software package (version 2.21.7, http://broadinstitute.github.io/picard/) following GATK Best Practices Guidelines[63]. Variant quality score recalibration (VQSR) was performed to exclude low-quality variant calls.

Additional quality control procedures were performed using VCFtools (version 0.1.16)[64], BCFtools (version 1.9)[65], PLINK (version 1.9)[66] and the Fratools PERL package (version 1, https://github.com/atks/fratools) at genotype, variant call and sample quality check levels. Low-quality genotype calls (read depth (DP) < 8 and genotype quality score (GQ) < 20) were excluded. Variants with low call rates in PD cases/controls/all samples (<95%), failed tests for Hardy–Weinberg equilibrium ($P < 1 \times 10^{-6}$) and high differential missingness between PD cases and controls ($P < 1 \times 10^{-4}$) were excluded from further analyses. We removed PD cases with sex discordance between genetically inferred sex from WES and patient records and samples showing poor genotyping concordance (<95% concordance) between WES data and genome-wide association arrays. Samples with excessive heterozygosity (>3.5 s.d. from mean), unusually high exome-wide singleton counts and low genotyping completion rate (<95%) were also excluded. We used identity-by-descent (IBD) analysis to identify related sample pairs up to the third degree (IBD > 0.125). For these pairs, the sample with the lower call rate was removed. Lastly, samples with outlying genetic ancestry as visualized using ancestry PCA with 2,141 samples from the 1000 Genomes Project reference panel[67] and 268 samples from the Singapore Genome Variation Project reference panel[68] were also removed. Variant calling and filtering was done across all samples equally regardless of affection status. All steps related to sample quality checks are summarized in Supplementary Table 3.

Identification of rare damaging genetic variants

Single-nucleotide variants and small insertions/deletions were annotated using Ensembl Variant Effect Predictor (VEP) software (version 104)[69]. Variant annotation was based on canonical transcripts defined in Ensembl VEP. Rare variants were defined based on ≤1% allele frequency in the WES dataset and the gnomAD East Asian population[50]. Functional

predictions for missense variants were derived from PolyPhen-2 (version 2.2.3r406)[70], and the VEP plugin LOFTEE (version 1.0.3)[71] was used for functional prediction of stop-gain/loss, splice site and frameshift variants. Rare variants predicted to be pathogenic (probably or possibly damaging) by PolyPhen-2/LOFTEE were termed 'qualifying' variants. Rare variants predicted as loss of function (LOFTEE 'high confidence' prediction) were termed PTVs. For the Replication datasets, published variant lists from whole-exome and whole-genome sequencing of *GBA1* and *SMPD1* were re-annotated based on the listed variant position in the same way as the Discovery dataset using PolyPhen-2/LOFTEE and Ensembl VEP software.

Validation by Sanger capillary sequencing

DNA samples from rare *SMPD1* allele carriers were polymerase chain reaction (PCR) amplified and capillary sequenced to confirm the rare allele calls from WES. PCR primers were designed using Primer3 (https://primer3.ut.ee/; Supplementary Tables 11 and 20).

ASM activity assay

To assess ASM activity associated with *SMPD1* variants observed, we generated HEK293T cell lines that contain individual *SMPD1* variants (Supplementary Table 21 and Supplementary Methods) and conducted a fluorescence-based assay to measure ASM activity.

We transduced HEK293T cell lines with individual *SMPD1* variants cloned into the lentiviral pLVX-CMV vector containing a puromycin selection marker[72] by standard lentiviral transduction methods. Fragments containing *SMPD1* variants were generated by one of the three approaches. In the first approach, *SMPD1* allele was generated from two separate overlapping fragments containing the variant-specific mutation. These two fragments were amplified from the open reading frame of the *SMPD1* canonical transcript ENST00000342245.4 using primer pairs (Supplementary Table 21). One fragment will be generated from amplification with the forward end primer (SMPD1_FWD) and variant-specific reverse primer, whereas the other fragment will be generated from amplification with the variant-specific forward primer and reverse end primer (SMPD1_REV). Both fragments and the linearized pLVX-CMV vector were joined by Gibson assembly (NEB, E2621S). In the second approach, a single fragment for *SMPD1* alleles close to the 3′ or 5′ ends of the *SMPD1* open reading frame was produced by incorporating the variant-specific mutation into the variant-specific primer by PCR. The single fragment and the linearized pLVX-CMV vector were joined by Gibson assembly (NEB, E2621S). In the third approach, fragments for *SMPD1* alleles that were technically challenging to generate with the PCR-based approach were synthesized using the GenBrick synthesis service from GenScript, before restriction digestion and ligation with the linearized pLVX-CMV vector. Sanger sequencing was conducted on all generated plasmids to confirm variant sequences before transduction into HEK293T cells.

Transduced cells underwent selection in puromycin-containing media for 4 d. Western blotting was conducted with anti-SMPD1 antibody (R&D Systems, MAB5348) to verify successful transduction by ascertainment of SMPD1 expression. A total of 126 cell lines expressing wild-type *SMPD1* and 125 individual protein-altering *SMPD1* genetic variants were made (122 rare variants and three common variants; Supplementary Table 17).

To measure ASM activity associated with each *SMPD1* genetic variant, cells were sonicated with protease inhibitor (Roche, 05056489001). Then, 10 µg of extract was tested for ASM activity in replicates ($n = 3–6$) using a fluorescence assay with synthetic substrate HMU-PC (Biosynth, EH31028)[73]. We measured the ASM activity of each *SMDP1* allele as the amount of fluorescent signal generated over 1 h at 37 °C. Measured activity for each variant was presented as percentage reduction in activity as compared to wild-type *SMPD1* that was calculated based on $\log_2$ fold change of signal per replicate by the average wild-type signal.

In silico structural analysis of p.Pro282Thr, p.Tyr500His and p.Asn177Tyr variants

Steric clashes due to p.Pro282Thr, p.Tyr500His and p.Asn177Tyr mutations were evaluated using the mutagenesis toolbox in PyMol (version 2.5.2) based on the atomic coordinates of human ASM bounded with zinc and phosphocholine (Protein Data Bank (PDB) ID: 5I85).

Expression of PD-associated genes in single-cell transcriptomics dataset

Single-nuclei transcriptomics data from four midbrain or substantia nigra studies, which consisted of 17 PD cases and 27 controls[74–77], were downloaded from the Gene Expression Omnibus (Supplementary Table 5).

Quality control and filtering were conducted with Seurat (version 4.3.0.1)[78] to retain cells with mitochondrial gene content less than 5%, 200–5,000 total genes expressed and singlets as identified with scDblFinder (version 1.15.2)[79]. A total of 117,338 single cells from 17 PD cases and 156,035 single cells from 27 controls remained after quality control and filtering. Gene expression measurements for retained cells were normalized with the SCTransform function and integrated. Dimensionality reduction by PCA and Uniform Manifold Approximation and Projection (UMAP) embedding was performed, followed by visualization of clustering stability with clustree (version 0.5.0). Cell clusters were annotated with known cell type markers for neuronal, glial, endothelial cell and immune cell types (Supplementary Table 7). The expression of PD-associated genes was assessed across the annotated cell types. Differential gene expression between PD and control for each cell type was assessed with the Seurat FindMarkers function using the Wilcoxon rank-sum test.

Postmortem midbrain samples

Snap-frozen human postmortem brain tissue samples were acquired from the Multiple Sclerosis and Parkinson's Tissue Bank (Imperial College London, Hammersmith Campus). Samples from the substantia nigra pars compacta (SNpc) region were obtained from two PD cases and one control (no PD clinical history and no PD-associated neuropathology). Written informed consent was obtained from each donor in strict adherence to the Declaration of Helsinki. All neuropathological diagnoses were made based on the BrainNet Europe protocol[80–82]. This research was approved by the Nanyang Technological University Institutional Review Board (IRB-2016-12-029 and IRB-2018-09-052) and the National Research Ethics Committee (08/MRE09/31 and National Health Service REC no. 18/SW/0029).

Protein expression of GBA1 and SMPD1 in brain tissues

Western blotting and immunohistochemistry were conducted for SMPD1 and GBA1 on snap-frozen postmortem SNpc sections from two PD cases and one control by lysis in RIPA buffer (Thermo Fisher Scientific) containing protease and phosphatase inhibitors (Thermo Fisher Scientific) on ice for 20 min, followed by 15-min centrifugation at 12,000*g* at 4 °C. Then, 5% 2-mercaptoethanol (Sigma-Aldrich) and 2× Laemmli sample buffer (Bio-Rad) were added to the protein lysates before heating at 95 °C for 5 min. Western blotting was then conducted on 35 µg of tissue lysate with anti-SMPD1 antibody (1:1,000 dilution; R&D Systems, MAB5348), anti-GBA1 antibody (1:100 dilution; Santa Cruz Biotechnology, sc-166407) and anti-GAPDH antibody (1:1;000 dilution; Bio-Rad, VMA00046).

Immunohistochemistry was conducted on snap-frozen postmortem SNpc tissues from the same two PD cases and one control by fixing snap-frozen tissue sections (10 µm thick) with 10% neutral buffered formalin for 30 min and cold methanol for 10 min. Blocking was conducted for 1 h with 2.5% normal goat serum (Vector Laboratories) before incubation with anti-SMPD1 antibody (1:500 dilution; R&D Systems, MAB5348), anti-GBA1 antibody (1:50 dilution; Santa Cruz Biotechnology, sc-166407), anti-GAPDH antibody

(1:1,000 dilution; Bio-Rad, VMA00046), anti-TH (1:200 dilution; Abcam, ab137869) and anti-HuC/D (1:500 dilution; Life Technologies, A-21271) overnight at 4 °C. Sections were incubated in 0.3% hydrogen peroxide for 10 min before adding ImmPRESS horseradish peroxidase polymer reagent (Vector Laboratories) for 1 h. The slides were subsequently developed with an ImmPACT-DAB substrate kit (Vector Laboratories) with Mayer's hematoxylin as counterstain. Sequential dehydration in 70%, 90% and 100% IMS was done, followed by clearing in 100% xylene. The sections were then mounted with DPX mountant (Sigma-Aldrich).

Immunostaining for SMPD1 co-localization with lysosomes

Immunostaining for SMPD1 was conducted on LysoTracker Red–stained HEK293T wild-type SMPD1 cells and HEK293T cells containing empty vector. HEK293T wild-type SMPD1 cells and HEK293T cells containing empty vector were seeded on Cell-Tak-coated coverslips (Corning, 354240) and cultured for 48 h. Cells were then treated with 5 nM LysoTracker Red DND-99 (Thermo Fisher Scientific, L7528) for 45 min and fixed with 4% (v/v) paraformaldehyde (Thermo Fisher Scientific, FB002) for 15 min at room temperature. Cells were permeabilized with 0.1% Triton X-100 (Sigma-Aldrich, T9284) for 10 min and blocked with 5% BSA (Sigma-Aldrich, SRE0036) for 1 h (both at room temperature). Incubation with anti-SMPD1 antibody (1:500 dilution; R&D Systems, MAB5348) was conducted overnight at 4 °C, followed by incubation with goat anti-mouse Alexa Fluor Plus 488 (1:500 dilution; Thermo Fisher Scientific, A32723) for 2 h at room temperature. DAPI counterstaining was conducted, and coverslips were mounted for confocal imaging on a Zeiss LSM 800 confocal laser scanning microscope, with image acquisition with the ×63 objective.

Statistics and reproducibility

Gene-based association testing. We tested whether any of the 17,396 autosomal genes across the human exome bear an excess or deficit of 'damaging' mutational burden in PD cases compared to controls. For the Discovery cohorts, a stratified Cochran–Mantel–Haenszel (CMH) test without continuity correction was used to evaluate gene-based burden across the exomes of participants from the five countries studied (Supplementary Table 2). Singapore and Malaysia (SG/MAL) samples were considered as one stratum due to the similarity in genetic background (as evidenced by PCA of SG/MAL samples; Supplementary Figs. 2 and 3). Haldane–Anscombe correction on OR was applied on carrier count matrices with at least one zero count[83,84]. We also evaluated gene-based burden with SKAT-O[85] after adjustment for per-sample variant counts, average per-sample coverage and principal components (PC1–PC3) as covariates.

Exome-wide significance was preset at $P \leq 2.5 \times 10^{-6}$ (two-tailed), based on multiple hypothesis testing correction for an estimated 20,000 protein-coding genes in the human genome. Genome-wide significance threshold was set at 5×10^{-8}, following the standard threshold in GWASs.

Replication of *SMPD1* and *GBA1* association. For the Replication cohorts, the same CMH test without continuity correction was conducted on a total of 5,585 PD cases and 5,642 controls from China, the United States and the IPDGC (Supplementary Table 16) in a stratified meta-analysis as performed for the Discovery cohorts. Gene-based testing of *SMPD1* qualifying variants was conducted across seven strata consisting of four Discovery cohorts (SG/MAL, HK, KR and TW) and three Replication cohorts (Zhao, Robak and Alcalay-Boston) with *SMPD1* data available. Gene-based testing of *GBA1* qualifying variants was conducted across six strata consisting of four Discovery cohorts (SG/MAL, HK, KR and TW) and two Replication cohorts (Zhao and Robak) with *GBA1* data available. Data are presented as OR and 95% CI. All *P* values presented are unadjusted.

Association of variant carrier status with PD onset. Association testing of *SMPD1*, *HLA-DRB1* and *GBA1* variant carrier status with age of PD onset was conducted with two-tailed *t*-tests on 4,149 of 4,298 PD cases that have available information on PD onset age.

Association of *SMPD1* variant carrier status with polygenic risk score. Association testing of *SMPD1* variant carrier status with polygenic risk score was also conducted using a two-tailed *t*-test. Polygenic risk score was based on a total of 67 genome-wide significant loci from both East Asian and European studies presented in our analysis. This included 11 loci reaching genome-wide significance in the East Asian GWAS[39] and 78 European risk loci from Nalls et al.[7]. At nine loci that overlap between East Asian and European, the East Asian risk SNP was used, as described in Foo et al.[39]. Of all 80 (78 + 11 − 9) unique SNPs that were considered, 67 were polymorphic and passed genotyping/imputation quality control filtering in the PD case and control samples analyzed in this study[39,86].

ASM activity assay. Cell lines bearing *SMPD1* variant alleles were assayed across different plates. To correct for plate-to-plate variation, a reference cell line expressing wild-type *SMPD1* alleles was always assayed together with the cell lines bearing *SMPD1* variant alleles. This allowed for direct comparison of fluorescent signal from *SMPD1* variant-containing cell extracts with that from the *SMPD1* wild-type reference cell line within each batch of experiment.

To assess passage-to-passage activity variation, we randomly selected eight cell lines containing eight different *SMPD1* variants and measured their resultant ASM activities using the same ASM activity assay in technical triplicates. The measured ASM activity of all eight variants relative to wild-type *SMPD1* remained highly concordant between passages (Supplementary Fig. 7).

Association of *SMPD1* variant ASM activity levels with domain localization. Rare *SMPD1* variants were classified by their localization to protein domains across the ASM protein and their effect on protein tertiary structure integrity based on protein crystallography studies (PDB: 5I81). Two-sided *t*-tests were used to compare ASM activity levels across each category.

Statistical analysis of ASM activity in Discovery and Replication cohorts. Because ASM enzymatic activity was measured as a continuous variable, we used iterative linear discriminant analysis to explore enzymatic activity thresholds in which variants below a tested threshold would be classified as 'impaired function'. This was done to obtain the best discrimination (in terms of *P* value) between PD cases and controls. Only functionally impaired variants with enzymatic activities below the defined threshold were included in the revised stratified CMH test for *SMPD1*.

Immunostaining for *SMPD1* co-localization with lysosomes. Each cell line was plated for immunostaining assay on coated coverslips in three wells across a six-well plate. After culturing and immunostaining procedures described above, one field of view was captured from each well, resulting in three fields of view captured per condition.

Protein expression and immunostaining of GBA1 and SMPD1 in brain tissues. The substantia nigra region was present in only 2–3 slices (10 µm) in each brain tissue block. As such, one slice per sample (two PD cases and one control) was lysed for western blotting for GBA1, SMPD1 and GAPDH protein expression, and one other slice per sample (two PD cases and one control) was used for immunohistochemistry for SMPD1, GBA1, TH and HuC/D.

Reporting Summary

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

Data availability

All unrestricted data are available from the corresponding author upon reasonable request. Summary statistics can be found in Supplementary Table 22. Western blots are provided as Supplementary Data. Acid sphingomyelinase (ASM) enzymatic activity assay absorbance readings are provided as Source Data. Individual-level genotype data are not publicly available because of consent and privacy issues but will be made available upon reasonable request to J.N.F. (jianee.foo@ntu.edu.sg; response timeframe of 1 month) and after approval by the relevant institutional review board. Midbrain single-nuclei RNA sequencing datasets used are available from the Gene Expression Omnibus as detailed in Supplementary Table 5. Protein crystallography of ASM tertiary structure is available from the Protein Data Bank (ID: 5I85).

Code availability

The codes used in this study can be found on GitHub (https://github.com/fjnlab/PD_exome-seq). Software and software packages used include the following: R (version 4.0.2) and R packages data.table (version 1.14.0), dplyr (version 1.0.5), tidyr (version 1.1.2), ggplot2 (version 3.3.6), forestplot (version 2.0.1), stats (mantelhaen.test, version 4.3.1) and SKAT (version 2.2.5); BWA (version 0.7.17); GATK (version 3.7); Picard (version 2.21.7); VCFtools (version 0.1.16); BCFtools (version 1.9); PLINK (version 1.9); Fratools PERL package (version 1, https://github.com/atks/fratools); Ensembl Variant Effect Predictor (VEP) software (version 104); ANNOVAR (version dated 2020-06-07); VEP plugin LOFTEE (version 1.0.3); PolyPhen-2 (version 2.2.3r406); dbNSFP (version 3.5a); Seurat (version 4.3.0.1); scDblFinder (version 1.15.2); clustree (version 0.5.0); and PyMol (version 2.5.2).

References

1. Nutt, J. G. & Wooten, G. F. Clinical practice. Diagnosis and initial management of Parkinson's disease. *N. Engl. J. Med.* **353**, 1021–1027 (2005).
2. Bloem, B. R., Okun, M. S. & Klein, C. Parkinson's disease. *Lancet* **397**, 2284–2303 (2021).
3. 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).
4. Simón-Sánchez, J. et al. Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat. Genet.* **41**, 1308–1312 (2009).
5. Nalls, M. A. et al. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet* **377**, 641–649, (2011).
6. Nalls, M. A. et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat. Genet.* **46**, 989–993 (2014).
7. 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).
8. Chang, D. et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat. Genet.* **49**, 1511–1516 (2017).
9. Kim, J. J. et al. Multi-ancestry genome-wide association meta-analysis of Parkinson's disease. *Nat. Genet.* **56**, 27–36 (2024).
10. Cohen, J. C., Boerwinkle, E., Mosley, T. H. Jr. & Hobbs, H. H. Sequence variations in *PCSK9*, low LDL, and protection against coronary heart disease. *N. Engl. J. Med.* **354**, 1264–1272 (2006).
11. Minikel, E. V. et al. Evaluating drug targets through human loss-of-function genetic variation. *Nature* **581**, 459–464 (2020).
12. Polymeropoulos, M. H. et al. Mapping of a gene for Parkinson's disease to chromosome 4q21-q23. *Science* **274**, 1197–1199 (1996).
13. Zimprich, A. et al. A mutation in *VPS35*, encoding a subunit of the retromer complex, causes late-onset Parkinson disease. *Am. J. Hum. Genet.* **89**, 168–175 (2011).
14. Vilariño-Güell, C. et al. *VPS35* mutations in Parkinson disease. *Am. J. Hum. Genet.* **89**, 162–167 (2011).
15. Kitada, T. et al. Mutations in the *parkin* gene cause autosomal recessive juvenile parkinsonism. *Nature* **392**, 605–608 (1998).
16. Bonifati, V. et al. Mutations in the *DJ-1* gene associated with autosomal recessive early-onset parkinsonism. *Science* **299**, 256–259 (2003).
17. Valente, E. M. et al. Hereditary early-onset Parkinson's disease caused by mutations in *PINK1*. *Science* **304**, 1158–1160 (2004).
18. Ramirez, A. et al. Hereditary parkinsonism with dementia is caused by mutations in *ATP13A2*, encoding a lysosomal type 5 P-type ATPase. *Nat. Genet.* **38**, 1184–1191 (2006).
19. Morgan, N. V. et al. *PLA2G6*, encoding a phospholipase A2, is mutated in neurodegenerative disorders with high brain iron. *Nat. Genet.* **38**, 752–754 (2006).
20. Di Fonzo, A. et al. *FBXO7* mutations cause autosomal recessive, early-onset parkinsonian-pyramidal syndrome. *Neurology* **72**, 240–245 (2009).
21. Quadri, M. et al. Mutation in the *SYNJ1* gene associated with autosomal recessive, early-onset Parkinsonism. *Hum. Mutat.* **34**, 1208–1215 (2013).
22. Deng, H. X. et al. Identification of *TMEM230* mutations in familial Parkinson's disease. *Nat. Genet.* **48**, 733–739 (2016).
23. Saini, P. et al. Association study of *DNAJC13*, *UCHL1*, *HTRA2*, *GIGYF2*, and *EIF4G1* with Parkinson's disease. *Neurobiol. Aging* **100**, 119.e7–119.e13 (2021).
24. Paisán-Ruíz, C. et al. Cloning of the gene containing mutations that cause *PARK8*-linked Parkinson's disease. *Neuron* **44**, 595–600 (2004).
25. Zimprich, A. et al. Mutations in *LRRK2* cause autosomal-dominant parkinsonism with pleomorphic pathology. *Neuron* **44**, 601–607 (2004).
26. Sidransky, E. et al. Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. *N. Engl. J. Med.* **361**, 1651–1661 (2009).
27. Makarious, M. B. et al. Large-scale rare variant burden testing in Parkinson's disease. *Brain* **146**, 4622–4632 (2023).
28. Zuk, O. et al. Searching for missing heritability: designing rare variant association studies. *Proc. Natl Acad. Sci. USA* **111**, E455–E464 (2014).
29. Correia Guedes, L., Mestre, T., Outeiro, T. F. & Ferreira, J. J. Are genetic and idiopathic forms of Parkinson's disease the same disease? *J. Neurochem.* **152**, 515–522 (2020).
30. Day, J. O. & Mullin, S. The genetics of Parkinson's disease and implications for clinical practice. *Genes (Basel)* **12**, 1006 (2021).
31. Klein, C. & Westenberger, A. Genetics of Parkinson's disease. *Cold Spring Harb. Perspect. Med.* **2**, a008888 (2012).
32. Di Lazzaro, G. et al. X-linked Parkinsonism: phenotypic and genetic heterogeneity. *Mov. Disord.* **36**, 1511–1525 (2021).
33. Zhao, Y.-W. et al. The association between lysosomal storage disorder genes and Parkinson's disease: a large cohort study in Chinese mainland population. *Front. Aging Neurosci.* **13**, 749109 (2021).
34. Robak, L. A. et al. Excessive burden of lysosomal storage disorder gene variants in Parkinson's disease. *Brain* **140**, 3191–3203 (2017).
35. Full spectrum genetics. *Nat. Genet.* **44**, 1 (2011).
36. Kornhuber, J., Rhein, C., Müller, C. P. & Mühle, C. Secretory sphingomyelinase in health and disease. *Biol. Chem.* **396**, 707–736 (2015).
37. Levran, O., Desnick, R. J. & Schuchman, E. H. Identification and expression of a common missense mutation (L302P) in the acid sphingomyelinase gene of Ashkenazi Jewish type A Niemann–Pick disease patients. *Blood* **80**, 2081–2087 (1992).

38. Levran, O., Desnick, R. J. & Schuchman, E. H. Niemann–Pick disease: a frequent missense mutation in the acid sphingomyelinase gene of Ashkenazi Jewish type A and B patients. *Proc. Natl Acad. Sci. USA* **88**, 3748–3752 (1991).
39. 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).
40. Le Guen, Y. et al. Multiancestry analysis of the HLA locus in Alzheimer's and Parkinson's diseases uncovers a shared adaptive immune response mediated by *HLA-DRB1*04* subtypes. *Proc. Natl Acad. Sci. USA* **120**, e2302720120 (2023).
41. Naito, T. et al. Trans-ethnic fine-mapping of the major histocompatibility complex region linked to Parkinson's disease. *Mov. Disord.* **36**, 1805–1814 (2021).
42. Kalogeropulou, A. F. et al. Impact of 100 LRRK2 variants linked to Parkinson's disease on kinase activity and microtubule binding. *Biochem. J* **479**, 1759–1783 (2022).
43. Wang, Q. et al. Rare variant contribution to human disease in 281,104 UK Biobank exomes. *Nature* **597**, 527–532 (2021).
44. Spargo, T. P. et al. Haploinsufficiency of *ITSN1* is associated with Parkinson's disease. Preprint at *medRxiv* https://doi.org/10.1101/2024.07.25.24310988 (2024).
45. Beutler, E. Gaucher disease: new molecular approaches to diagnosis and treatment. *Science* **256**, 794–799 (1992).
46. Schuchman, E. H. The pathogenesis and treatment of acid sphingomyelinase-deficient Niemann–Pick disease. *J. Inherit. Metab. Dis.* **30**, 654–663 (2007).
47. Lim, J. L. et al. *Glucocerebrosidase* (*GBA*) gene variants in a multi-ethnic Asian cohort with Parkinson's disease: mutational spectrum and clinical features. *J. Neural Transm. (Vienna)* **129**, 37–48 (2022).
48. Gan-Or, Z. et al. The p.L302P mutation in the lysosomal enzyme gene *SMPD1* is a risk factor for Parkinson disease. *Neurology* **80**, 1606–1610 (2013).
49. Wu, R. M., Lin, C. H. & Lin, H. I. The p.L302P mutation in the lysosomal enzyme gene *SMPD1* is a risk factor for Parkinson disease. *Neurology* **82**, 283 (2014).
50. Chen, S. et al. A genomic mutational constraint map using variation in 76,156 human genomes. *Nature* **625**, 92–100 (2024).
51. Foo, J.-N. et al. A rare lysosomal enzyme gene *SMPD1* variant (p.R591C) associates with Parkinson's disease. *Neurobiol. Aging* **34**, 2890.e2813–2890.e2815 (2013).
52. Gan-Or, Z. et al. The emerging role of *SMPD1* mutations in Parkinson's disease: implications for future studies. *Parkinsonism Relat. Disord.* **21**, 1294–1295 (2015).
53. Chen, Y. P. et al. Rare variants analysis of lysosomal related genes in early-onset and familial Parkinson's disease in a Chinese cohort. *J. Parkinsons Dis.* **11**, 1845–1855 (2021).
54. Mao, C. Y. et al. *SMPD1* variants in Chinese Han patients with sporadic Parkinson's disease. *Parkinsonism Relat. Disord.* **34**, 59–61 (2017).
55. Ylönen, S. et al. Genetic risk factors in Finnish patients with Parkinson's disease. *Parkinsonism Relat. Disord.* **45**, 39–43 (2017).
56. Abbott, S. K. et al. Altered ceramide acyl chain length and ceramide synthase gene expression in Parkinson's disease. *Mov. Disord.* **29**, 518–526 (2014).
57. Xicoy, H., Brouwers, J. F., Wieringa, B. & Martens, G. J. M. Explorative combined lipid and transcriptomic profiling of substantia nigra and putamen in Parkinson's disease. *Cells* **9**, 1966 (2020).
58. Alcalay, R. N. et al. *SMPD1* mutations, activity, and α-synuclein accumulation in Parkinson's disease. *Mov. Disord.* **34**, 526–535 (2019).
59. Rhein, C. et al. The acid sphingomyelinase sequence variant p.A487V is not associated with decreased levels of enzymatic activity. *JIMD Rep.* **8**, 1–6 (2013).
60. Li, Z. et al. Association of rare *CYP39A1* variants with exfoliation syndrome involving the anterior chamber of the eye. *JAMA* **325**, 753–764 (2021).
61. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at *arXiv* https://doi.org/10.48550/arXiv.1303.3997 (2013).
62. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
63. Van der Auwera, G. & O'Connor, B. *Genomics in the Cloud* (O'Reilly Media, Inc., 2020).
64. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
65. 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).
66. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
67. Delaneau, O. et al. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat. Commun.* **5**, 3934 (2014).
68. Teo, Y. Y. et al. Singapore Genome Variation Project: a haplotype map of three Southeast Asian populations. *Genome Res.* **19**, 2154–2162 (2009).
69. McLaren, W. et al. The Ensembl Variant Effect Predictor. *Genome Biol.* **17**, 122 (2016).
70. Adzhubei, I. A. et al. A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
71. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
72. Wang, Z. et al. Methionine is a metabolic dependency of tumor-initiating cells. *Nat. Med.* **25**, 825–837 (2019).
73. van Diggelen, O. P. et al. A new fluorimetric enzyme assay for the diagnosis of Niemann–Pick A/B, with specificity of natural sphingomyelinase substrate. *J. Inherit. Metab. Dis.* **28**, 733–741 (2005).
74. Lee, A. J. et al. Characterization of altered molecular mechanisms in Parkinson's disease through cell type-resolved multiomics analyses. *Sci. Adv.* **9**, eabo2467 (2023).
75. Smajić, S. et al. Single-cell sequencing of human midbrain reveals glial activation and a Parkinson-specific neuronal state. *Brain* **145**, 964–978 (2022).
76. Agarwal, D. et al. A single-cell atlas of the human substantia nigra reveals cell-specific pathways associated with neurological disorders. *Nat. Commun.* **11**, 4183 (2020).
77. Kamath, T. et al. Single-cell genomic profiling of human dopamine neurons identifies a population that selectively degenerates in Parkinson's disease. *Nat. Neurosci.* **25**, 588–595 (2022).
78. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 (2021).
79. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329–337 (2019).
80. Alafuzoff, I. et al. Staging of neurofibrillary pathology in Alzheimer's disease: a study of the BrainNet Europe Consortium. *Brain Pathol.* **18**, 484–496 (2008).
81. Alafuzoff, I. et al. Assessment of β-amyloid deposits in human brain: a study of the BrainNet Europe Consortium. *Acta Neuropathol.* **117**, 309–320 (2009).

82. Attems, J. et al. Neuropathological consensus criteria for the evaluation of Lewy pathology in post-mortem brains: a multi-centre study. *Acta Neuropathol.* **141**, 159–172 (2021).
83. Anscombe, F. J. On estimating binomial response relations. *Biometrika* **43**, 461–464 (1956).
84. Haldane, J. B. S. The mean and variance of χ^2, when used as a test of homogeneity, when expectations are small. *Biometrika* **31**, 346–355 (1940).
85. Lee, S. et al. Optimal unified approach for rare-variant association testing with application to small-sample case–control whole-exome sequencing studies. *Am. J. Hum. Genet.* **91**, 224–237 (2012).
86. Khor, C. C. et al. Genome-wide association study identifies five new susceptibility loci for primary angle closure glaucoma. *Nat. Genet.* **48**, 556–562 (2016).

Acknowledgements

We thank W.-Y. Meah (Genome Institute of Singapore), S. H. Ng (National Neuroscience Institute), W. L. Beh, W. Tay, K. Wong and A. Ang (Nanyang Technological University Singapore) for administrative support. We thank S. W. Lim and Y. W. Y. Oo for assistance in structural modeling. We also thank P. C. Ng and H. Liany for their work on the early version of the variant annotation and analysis pipeline. This work is supported by the Singapore Ministry of Health's National Medical Research Council Open Fund Large Collaborative Grant (MOH-000207 to E.-K.T.; MOH-001072 to T.A. and C.C.K.); the Open Fund Individual Research Grant (MOH-000559 to J.N.F.; MOH-001110 to Z.W. and Y.S.H.; MOH-001214 to Y.X.); the Singapore Translational Research Investigator Award (STaR) (MOH-000435 to. T.A. and Z.W.); the Open Fund Young Individual Research Grant (MOH-001329 to E.GY.C); and the Singapore Ministry of Education Academic Research Fund Tier 2 (MOE-T2EP30220-0005 to J.N.F.; MOE-T2EP30220-0008 to Y.X.) and Tier 3 (MOE-MOET32020-0004 to J.N.F. and Y.X.). C.C.K. is supported by the Singapore National Research Foundation Investigatorship (NRF-NRFI2018-01). S.-Y.L. and A.H.T. are supported by the University of Malaya Parkinson's Disease and Movement Disorders Research Program (PV035-2017). Y.S.H. and E.KL.P. are supported by the Agency for Science, Technology and Research (A*STAR). R.N.A.'s research is funded by the Michael J. Fox Foundation, the Silverstein Foundation and the Parkinson's Foundation. Collection of DNA samples by J.L.W. was funded by National Institutes of Health (NIH) EY015473 and NIH P30 EY014104. Work by M.A.N. and C.B. was supported, in part, by the Intramural Research Program of the NIH, National Institute on Aging, NIH, Department of Health and Human Services (project no. ZO1 AG000534); the National Institute of Neurological Disorders and Stroke; the Office of Intramural Research; and the NIH Office of the Director; and used the computational resources of the NIH STRIDES Initiative (https://cloud.nih.gov) through the Other Transaction agreement: Azure: OT2OD032100, Google Cloud Platform: OT2OD027060, Amazon Web Services: OT2OD027852 and the NIH HPC Biowulf cluster (https://hpc.nih.gov). This work was also supported by the A*STAR Computational Resource Center through the use of its high-performance computing facilities.

Author contributions

The study was conceptualized by E.-K.T., C.C.K., Z.W. and J.N.F. Sample collection and patient assessments were conducted by S.J.C., E.Y.N., L.CS.T., L.L.C., Y.C., W.-L.A., K.M.P., J.L.L., Y.W.T., V.M., A.YY.C., J.-J.L., B.S.J., A.H.T., A.A.A., S.-Y.L. and E.-K.T. K.S., C.C.T., C.P.P., J.A., K.H.P., J.L.W., T.A., M.B.M., C.B., M.A.N., L.A.R., R.N.A. and Z.G.-O. contributed data. Experiments were designed by E.GY.C., Z.L., M.T., Y.J.H., W.L.C., T.J.T., E.KL.P., Y.S.H., C.H.C., T.X.P., R.R., Y.X., C.C.K., E.-K.T., Z.W. and J.N.F. Experiments were performed by E.GY.C., Z. Liu, M.T., Y.J.H., W.L.C., T.J.T., E.KL.P., X.Y.C., E.YT.L., C.H.C., J.-J.L. and T.X.P. Data analysis was conducted by E.GY.C., Z. Liu, Z. Li, M.M.L., M.T., Y.J.H., W.L.C., T.J.T., E.KL.P., Y.S.H., X.Y.C., E.YT.L., C.H.C., J.-J.L., T.X.P., J.L.W., L.A.R., R.R., Y.X., C.C.K., Z.W. and J.N.F. E.GY.C., C.C.K., Z.W. and J.N.F drafted the paper, with input from all other authors. All authors critically assessed the paper.

Competing interests

Z.G.-O. received consultancy fees from Lysosomal Therapeutics, Idorsia, Prevail Therapeutics, Ono Therapeutics, Denali, Handl Therapeutics, Neuron23, Bial Biotech, Bial, UCB, Capsida, Vanqua Bio, Congruence Therapeutics, Takeda, Jazz Pharmaceuticals, Guidepoint, Lighthouse and Deerfield. R.N.A. received consultation fees from Biogen, Biohaven, Capsida, Gain Therapeutics, Genzyme/Sanofi, Janssen, Servier, SK Biopharmaceuticals, Takeda and Vanqua Bio. Y.X. holds a stock option in NeoCytogen Therapeutics where she is scientific co-founder and Chief Scientific Officer. M.A.N.'s participation in this project was part of a competitive contract awarded to DataTecnica, LLC by the National Institutes of Health to support open science research. M.A.N. also owns stock in Character Bio, Inc. and Neuron23, Inc. The other authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s43587-024-00760-7.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43587-024-00760-7.

Correspondence and requests for materials should be addressed to Chiea Chuen Khor, Eng-King Tan, Zhenxun Wang or Jia Nee Foo.

Peer review information *Nature Aging* thanks Huw Morris and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

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

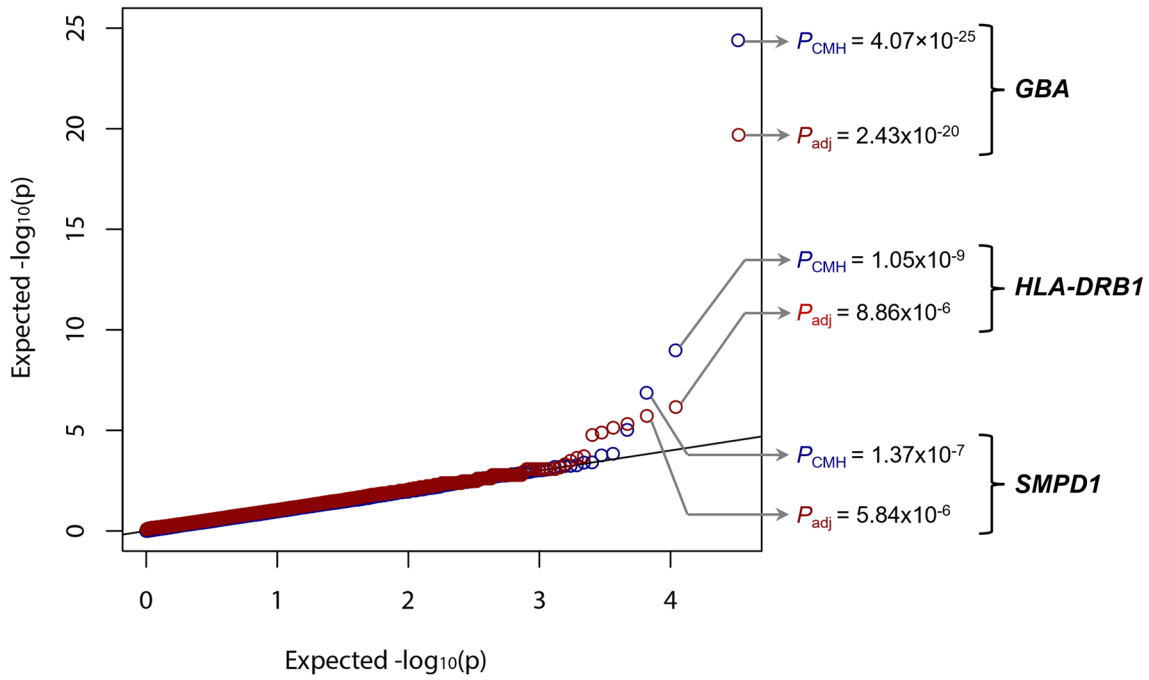
Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2024

Elaine GY Chew[1,34], Zhehao Liu[2,3,34], Zheng Li[3,34], Sun Ju Chung[4,34], Michelle M. Lian[1], Moses Tandiono[1], Yue Jing Heng[1], Ebonne Y. Ng[5], Louis CS Tan[6], Wee Ling Chng[2], Tiak Ju Tan[3], Esther KL Peh[7], Ying Swan Ho[7], Xiao Yin Chen[3], Erin YT Lim[3], Chu Hua Chang[1], Jonavan J. Leong[1], Ting Xuan Peh[3], Ling Ling Chan[2,8], Yinxia Chao[2,5], Wing-Lok Au[6], Kumar M. Prakash[2,5], Jia Lun Lim[9], Yi Wen Tay[9], Vincent Mok[10,11], Anne YY Chan[10], Juei-Jueng Lin[12], Beom S. Jeon[13], Kyuyoung Song[14], Clement C. Tham[15], Chi Pui Pang[15], Jeeyun Ahn[16,17], Kyu Hyung Park[17], Janey L. Wiggs[18], Tin Aung[2,19], Ai Huey Tan[20], Azlina Ahmad Annuar[9], Mary B. Makarious[21,22,23], Cornelis Blauwendraat[21,24], Mike A. Nalls[24,25], Laurie A. Robak[26,27], Roy N. Alcalay[28,29], Ziv Gan-Or[30,31,32], Richard Reynolds[1,33], Shen-Yang Lim[20], Yun Xia[1], Chiea Chuen Khor[2,3,19,35] ✉, Eng-King Tan[2,5,35] ✉, Zhenxun Wang[2,3,35] ✉ & Jia Nee Foo[1,3,35] ✉

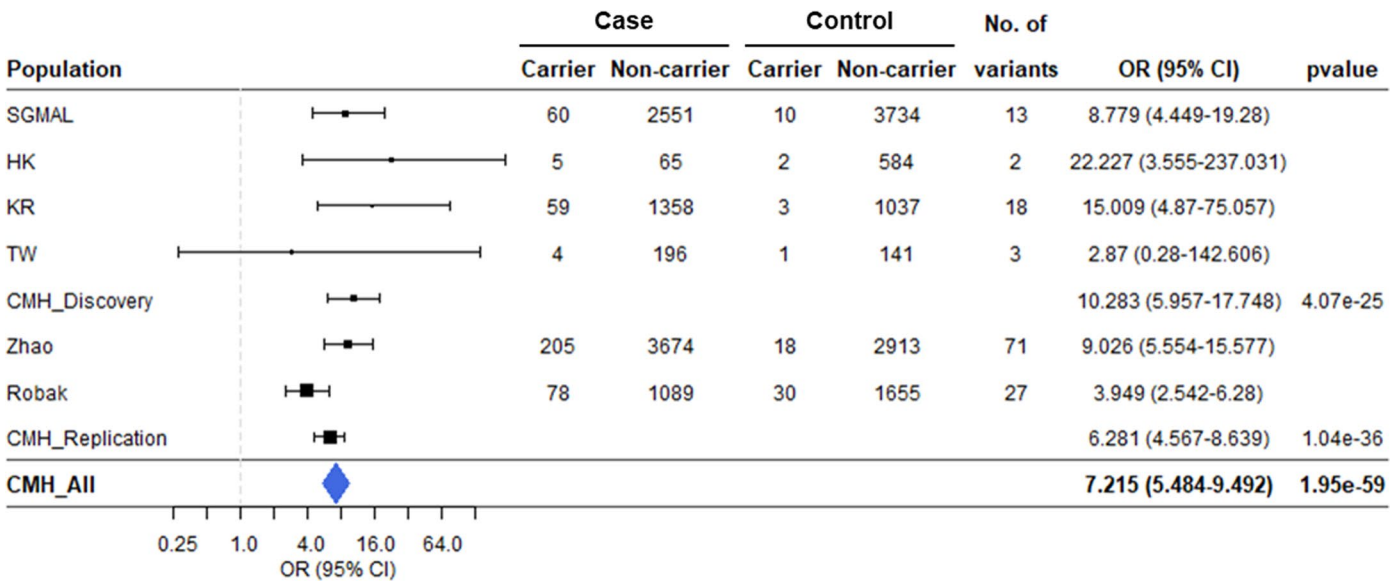
[1]Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore, Singapore. [2]Duke-National University of Singapore Medical School, Singapore, Singapore. [3]Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Singapore. [4]Department of Neurology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea. [5]Department of Neurology, National Neuroscience Institute, Singapore General Hospital, Singapore, Singapore. [6]Department of Neurology, National Neuroscience Institute, Singapore, Singapore. [7]Bioprocessing Technology Institute, Agency for Science, Technology and Research, Singapore, Singapore. [8]Department of Neuroradiology, Singapore General Hospital, Singapore, Singapore. [9]Department of Biomedical Science, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia. [10]Department of Medicine and Therapeutics, Division of Neurology, Margaret K.L. Cheung Research Centre for Management of Parkinsonism, Lui Che Woo Institute of Innovative Medicine, Prince of Wales Hospital, The Chinese University of Hong Kong, Hong Kong, China. [11]Gerald Choa Neuroscience Institute, Li Ka Shing Institute of Health Sciences, Hong Kong, China. [12]Department of Neurology, Chushang Show-Chwan Hospital, Nantou, Taiwan. [13]Department of Neurology, Seoul National University Hospital, Seoul, South Korea. [14]Department of Biochemistry and Molecular Biology, University of Ulsan College of Medicine, Seoul, South Korea. [15]Department of Ophthalmology and Visual Sciences, The Chinese University of Hong Kong, Hong Kong, China. [16]Department of Ophthalmology, Seoul Metropolitan Government, Seoul National University Boramae Medical Center, Seoul, South Korea. [17]Department of Ophthalmology, Seoul National University Hospital, Seoul, Korea. [18]Department of Ophthalmology, Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, MA, USA. [19]Singapore Eye Research Institute, Singapore, Singapore. [20]Division of Neurology, Department of Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia. [21]Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA. [22]Department of Clinical and Movement Neurosciences, UCL Queen Square Institute of Neurology, London, UK. [23]UCL Movement Disorders Centre, University College London, London, UK. [24]Center for Alzheimer's and Related Dementias, National Institute on Aging, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA. [25]Data Tecnica International, LLC, Bethesda, MD, USA. [26]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA. [27]Jan and Dan Duncan Neurologic Research Institute, Texas Children's Hospital, Houston, TX, USA. [28]Tel Aviv Sourasky Medical Center, Tel Aviv, Israel. [29]Columbia University Irving Medical Center, New York, NY, USA. [30]The Neuro (Montréal Neurological Institute-Hospital), McGill University, Montréal, QC, Canada. [31]Department of Human Genetics, McGill University, Montréal, QC, Canada. [32]Department of Neurology and Neurosurgery, McGill University, Montréal, QC, Canada. [33]Department of Brain Sciences, Faculty of Medicine, Imperial College London, London, UK. [34]These authors contributed equally: Elaine GY Chew, Zhehao Liu, Zheng Li, Sun Ju Chung. [35]These authors jointly supervised this work: Chiea Chuen Khor, Eng-King Tan, Zhenxun Wang, Jia Nee Foo. ✉e-mail: khorcc@gis.a-star.edu.sg; gnrtek@sgh.com.sg; zhenxun.wang@duke-nus.edu.sg; jianee.foo@ntu.edu.sg

https://doi.org/10.1038/s43587-024-00760-7



Extended Data Fig. 1 | Quantile-Quantile plot of all autosomal genes from the Discovery dataset. A stratified CMH test without continuity correction was used to evaluate gene-based burden across the exomes of the study participants from the 5 countries studied ("P_{CMH}") and indicated by blue circles. SKAT-O gene-based burden after adjustment for principal components (PC1-PC3), per sample variant counts and average per sample coverage ("P_{adj}") indicated by red circles. Details in Supplementary Table 22.

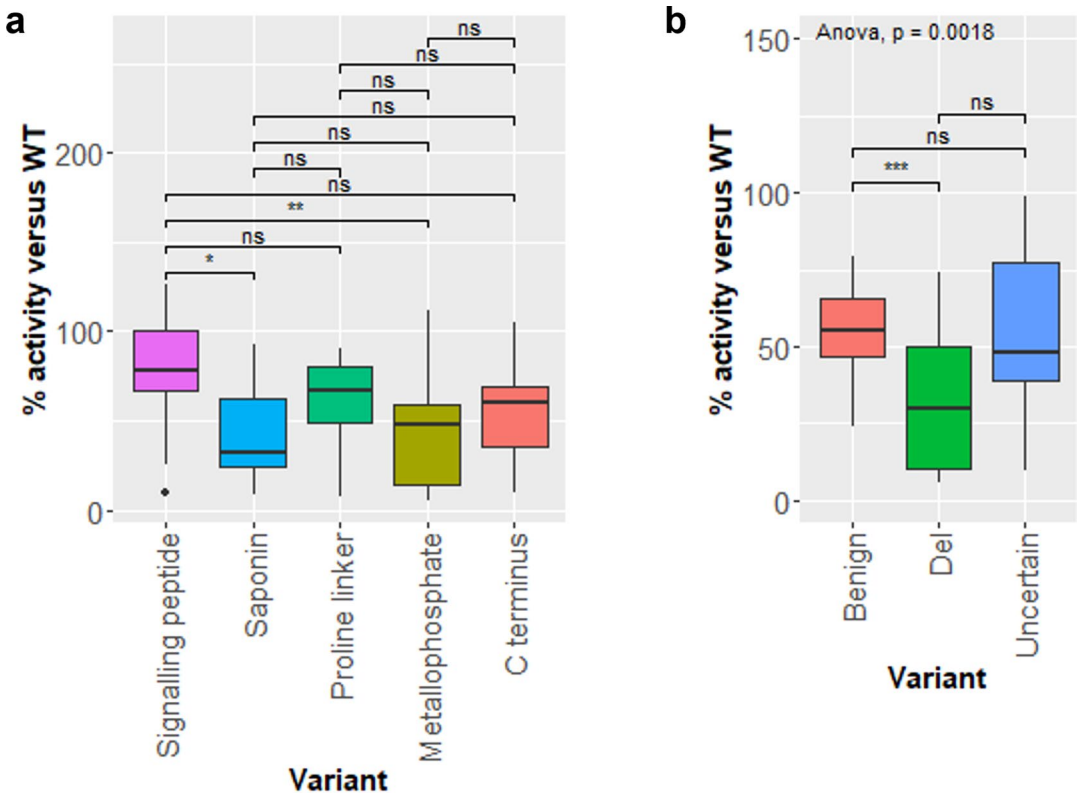
https://doi.org/10.1038/s43587-024-00760-7



Population	Case		Control		No. of		
	Carrier	Non-carrier	Carrier	Non-carrier	variants	OR (95% CI)	pvalue
SGMAL	60	2551	10	3734	13	8.779 (4.449-19.28)	
HK	5	65	2	584	2	22.227 (3.555-237.031)	
KR	59	1358	3	1037	18	15.009 (4.87-75.057)	
TW	4	196	1	141	3	2.87 (0.28-142.606)	
CMH_Discovery						10.283 (5.957-17.748)	4.07e-25
Zhao	205	3674	18	2913	71	9.026 (5.554-15.577)	
Robak	78	1089	30	1655	27	3.949 (2.542-6.28)	
CMH_Replication						6.281 (4.567-8.639)	1.04e-36
CMH_All						**7.215 (5.484-9.492)**	**1.95e-59**

Extended Data Fig. 2 | Replication of *GBA1* association in Chinese and European datasets based on rare predicted pathogenic variants. CMH without continuity correction used across strata. Data presented as OR and 95% confidence interval (CI).

https://doi.org/10.1038/s43587-024-00760-7



Extended Data Fig. 3 | Distribution of ASM enzymatic activity levels of rare *SMPD1* variants in relation to protein domain localisation and tertiary structure integrity. (**a**) Each rare *SMPD1* variant was classified by its localization to protein domains across the ASM protein, namely the signalling peptide (12 variants), saponin (12 variants), proline linker (7 variants), metallophosphate (70 variants), and C terminus (19 variants). (**b**) Each rare *SMPD1* variant was classified by its effect on ASM protein tertiary structure integrity as benign (15 variants), deleterious (27 variants) and uncertain (7 variants). Analysis for both panels are based on protein crystallography of ASM tertiary structure (PDB: 5i81). Median, 25th-75th quartiles, whiskers of 1.5× IQR shown in boxplot. Two-sided *t*-test: $P < 0.05$ indicated by *, $P < 0.01$ indicated by **; $P < 0.001$ indicated by ***.

https://doi.org/10.1038/s43587-024-00760-7

a Affects tertiary structure

Population	Case Carrier	Case Non-carrier	Control Carrier	Control Non-carrier	No. of variants	OR (95% CI)	pvalue
SGMAL	66	2545	42	3702	6	2.286 (1.524-3.462)	
HK	2	68	10	576	2	1.692 (0.177-8.186)	
KR	7	1410	2	1038	3	2.576 (0.489-25.461)	
TW	0	200	1	141	1	0 (0-27.69)	
CMH_Discovery						2.203 (1.531-3.171)	1.31e-05
Zhao	92	3787	39	2892	12	1.801 (1.222-2.7)	
Robak	36	1131	44	1641	10	1.187 (0.737-1.9)	
AlcalayBoston	13	537	28	1004	9	0.868 (0.409-1.749)	
CMH_Replication						1.399 (1.079-1.813)	1.18e-02
CMH_All						**1.63 (1.32-2.012)**	**5.02e-06**



b Signalling peptide

Population	Case Carrier	Case Non-carrier	Control Carrier	Control Non-carrier	No. of variants	OR (95% CI)	pvalue
SGMAL	2	2609	2	3742	4	1.434 (0.104-19.794)	
HK	0	70	0	586	0	0 (0-Inf)	
KR	3	1414	4	1036	3	0.55 (0.08-3.256)	
TW	1	199	0	142	1	Inf (0.018-Inf)	
CMH_Discovery						0.915 (0.295-2.842)	8.77e-01
Zhao	3	3876	0	2931	3	Inf (0.312-Inf)	
Robak	0	1167	1	1684	1	0 (0-56.262)	
AlcalayBoston	4	546	5	1027	2	1.504 (0.297-7.022)	
CMH_Replication						1.821 (0.581-5.708)	3.12e-01
CMH_All						**1.283 (0.575-2.862)**	**5.41e-01**



c Saponin

Population	Case Carrier	Case Non-carrier	Control Carrier	Control Non-carrier	No. of variants	OR (95% CI)	pvalue
SGMAL	0	2611	0	3744	0	0 (0-Inf)	
HK	0	70	0	586	0	0 (0-Inf)	
KR	0	1417	0	1040	0	0 (0-Inf)	
TW	0	200	0	142	0	0 (0-Inf)	
CMH_Discovery						NaN (NaN-NaN)	NaN
Zhao	18	3861	14	2917	10	0.971 (0.456-2.114)	
Robak	0	1167	1	1684	1	0 (0-56.262)	
AlcalayBoston	2	548	2	1030	1	1.879 (0.136-25.988)	
CMH_Replication						0.997 (0.519-1.914)	9.93e-01
CMH_All						**0.997 (0.519-1.914)**	**9.93e-01**



d Proline linker

Population	Case Carrier	Case Non-carrier	Control Carrier	Control Non-carrier	No. of variants	OR (95% CI)	pvalue
SGMAL	0	2611	0	3744	0	0 (0-Inf)	
HK	0	70	0	586	0	0 (0-Inf)	
KR	0	1417	0	1040	0	0 (0-Inf)	
TW	0	200	0	142	0	0 (0-Inf)	
CMH_Discovery						NaN (NaN-NaN)	NaN
Zhao	4	3875	15	2916	7	0.201 (0.048-0.631)	
Robak	0	1167	0	1685	0	0 (0-Inf)	
AlcalayBoston	0	550	0	1032	0	0 (0-Inf)	
CMH_Replication						0.201 (0.067-0.605)	1.55e-03
CMH_All						**0.201 (0.067-0.605)**	**1.55e-03**



e Metallophosphate

Population	Case Carrier	Case Non-carrier	Control Carrier	Control Non-carrier	No. of variants	OR (95% CI)	pvalue
SGMAL	83	2528	52	3692	20	2.331 (1.622-3.376)	
HK	3	67	12	574	5	2.139 (0.378-8.203)	
KR	16	1401	3	1037	11	3.946 (1.125-21.185)	
TW	0	200	1	141	1	0 (0-27.69)	
CMH_Discovery						2.379 (1.722-3.287)	6.66e-08
Zhao	110	3769	57	2874	39	1.471 (1.055-2.071)	
Robak	100	1067	66	1619	17	2.298 (1.651-3.218)	
AlcalayBoston	14	536	30	1002	12	0.872 (0.424-1.714)	
CMH_Replication						1.688 (1.364-2.088)	1.05e-06
CMH_All						**1.872 (1.568-2.236)**	**1.78e-12**



f C terminus

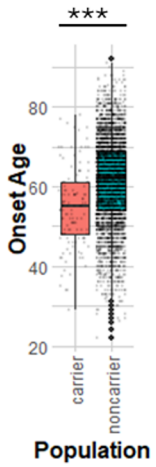
Population	Case Carrier	Case Non-carrier	Control Carrier	Control Non-carrier	No. of variants	OR (95% CI)	pvalue
SGMAL	16	2595	19	3725	10	1.209 (0.58-2.487)	
HK	0	70	0	586	0	0 (0-Inf)	
KR	2	1415	1	1039	2	1.468 (0.076-86.681)	
TW	0	200	1	141	1	0 (0-27.69)	
CMH_Discovery						1.146 (0.609-2.156)	6.70e-01
Zhao	18	3861	11	2920	9	1.238 (0.553-2.904)	
Robak	0	1167	2	1683	2	0 (0-7.688)	
AlcalayBoston	2	548	2	1030	4	1.879 (0.136-25.988)	
CMH_Replication						1.164 (0.589-2.299)	6.62e-01
CMH_All						**1.155 (0.727-1.835)**	**5.42e-01**



Extended Data Fig. 4 | Forest plot of PD associations with rare *SMPD1* variants from Discovery and Replication datasets that (a) affect ASM tertiary structure, (b) localized to the signaling peptide, (c) localized to the saponin domain, (d) localized to the proline-rich linker, (e) localized to the catalytic metallophosphate domain, and (f) localized to the C terminus. Effect of variants on ASM tertiary structure and localisation based on protein crystallography of ASM tertiary structure (PDB: 5I85). CMH without continuity correction used across strata. Data presented as OR and 95% confidence interval (CI).

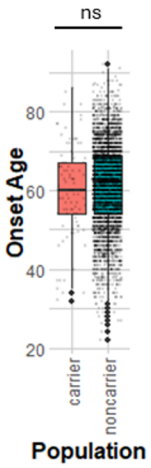
https://doi.org/10.1038/s43587-024-00760-7

a ***GBA1***



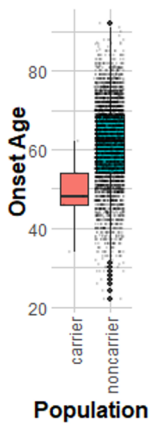
		No. of individuals	Mean	SD	Median	Min	Max	Two-sided t-test P
SGMAL	carrier	55	57.7	10.9	58.0	34.0	78.0	0.001
	non-carrier	2406	62.7	10.5	64.0	22.0	92.0	
HK	carrier	5	50.6	4.7	51.0	46.0	58.0	0.030
	non-carrier	64	57.2	10.7	58.0	24.0	78.0	
KR	carrier	59	51.9	8.7	52.0	29.0	74.0	8.38E-07
	non-carrier	1353	58.3	10.1	59.0	22.0	88.0	
TW	carrier	4	58.8	9.2	60.5	47.0	67.0	0.157
	non-carrier	196	67.3	10.5	69.5	29.0	91.0	
All	carrier	123	54.7	10.0	55.0	29.0	78.0	3.32E-11
	non-carrier	4019	61.4	10.7	62.0	22.0	92.0	

b ***SMPD1***



		No. of individuals	Mean	SD	Median	Min	Max	Two-sided t-test P
SGMAL	carrier	78	61.3	11.4	60.5	32.0	86.0	0.295
	non-carrier	2383	62.6	10.5	63.4	22.0	92.0	
HK	carrier	3	42.7	6.7	46.0	35.0	47.0	0.050
	non-carrier	66	57.4	10.3	58.0	24.0	78.0	
KR	carrier	12	57.8	12.0	60.0	37.0	73.0	0.932
	non-carrier	1400	58.1	10.2	59.0	22.0	88.0	
TW	carrier	NA	NA	NA	NA	NA	NA	NA
	non-carrier	200	67.2	10.5	69.0	29.0	91.0	
All	carrier	93	60.2	11.8	60.0	32.0	86.0	0.426
	non-carrier	4049	61.2	10.7	62.0	22.0	92.0	

c ***GBA1 + SMPD1***



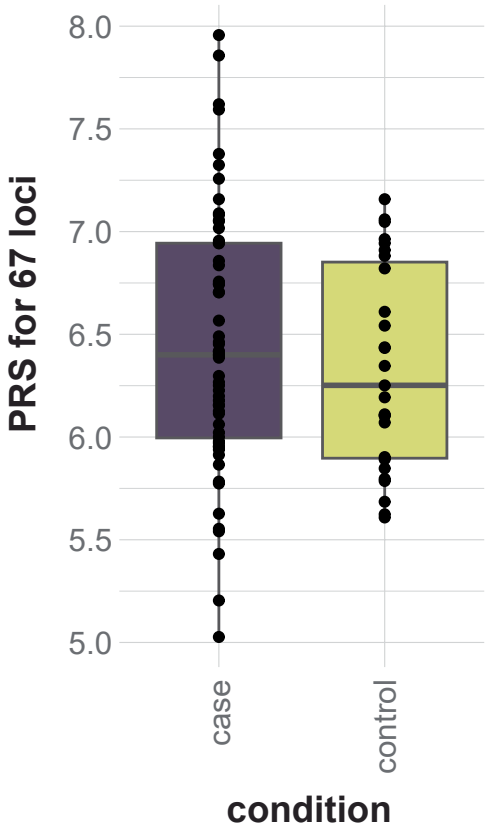
		No. of individuals	Mean	SD	Median	Min	Max	Two-sided t-test P
SGMAL	carrier	3	48.0	14.0	48.0	34.0	62.0	0.212
	non-carrier	2458	62.6	10.6	63.0	22.0	92.0	
HK	carrier	1	46.0	NA	46.0	46.0	46.0	NA
	non-carrier	68	56.9	10.5	57.5	24.0	78.0	
KR	carrier	1	54.0	NA	54.0	54.0	54.0	NA
	non-carrier	1411	58.1	10.2	59.0	22.0	88.0	
TW	carrier	NA	NA	NA	NA	NA	NA	NA
	non-carrier	200	67.2	10.5	69.0	29.0	91.0	
All	carrier	5	48.8	10.4	48.0	34.0	62.0	0.055
	non-carrier	4137	61.2	10.7	62.0	22.0	92.0	

Extended Data Fig. 5 | See next page for caption.

https://doi.org/10.1038/s43587-024-00760-7

Extended Data Fig. 5 | Onset ages of Parkinson's disease cases that carry rare deleterious variants in *GBA1* and *SMPD1*. (**a**) Carriers of *GBA1* qualifying variants were diagnosed with Parkinson's disease almost 7 years younger than non-carriers (carriers: median = 55.0 years of age, range = 29-78 years of age; non-carriers: median = 62.0 years of age, range = 22-92 years of age; *t*-test $P = 3.32 \times 10^{-11}$). (**b**) Carriers of *SMPD1* qualifying variants had similar onset age as non-carriers (carriers: median = 60.0 years of age, range = 32-86 years of age; non-carriers: median = 62.0 years of age, range = 22-92 years of age; *t*-test $P = 0.426$). (**c**) Onset ages in carriers of both *GBA1* and *SMPD1* qualifying variants. Median, 25th-75th quartiles, whiskers of 1.5× IQR shown in boxplot. *** indicates two-sided *t*-test $P = 3.32 \times 10^{-11}$.

https://doi.org/10.1038/s43587-024-00760-7

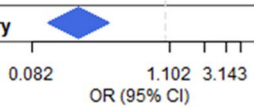


Extended Data Fig. 6 | Polygenic risk scores of individuals based on *SMPD1* functionally deficient variant carrier status. Polygenic risk scores of 60 cases and 27 controls which carry *SMPD1* functionally deficient variant shown. Median, 25th-75th quartiles, whiskers of 1.5× IQR shown in boxplot.

https://doi.org/10.1038/s43587-024-00760-7

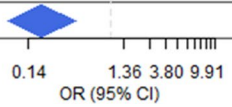
a

Population	Case		Control		No. of variants	OR (95% CI)	pvalue
	Carrier	Non-carrier	Carrier	Non-carrier			
SGMAL	4	2607	38	3706	6	0.15 (0.039-0.416)	
HK	0	70	2	584	2	0 (0-44.787)	
KR	10	1407	31	1009	4	0.231 (0.101-0.487)	
TW	1	199	2	140	2	0.353 (0.006-6.839)	
CMH_Discovery						**0.197 (0.111-0.35)**	**1.05e-09**



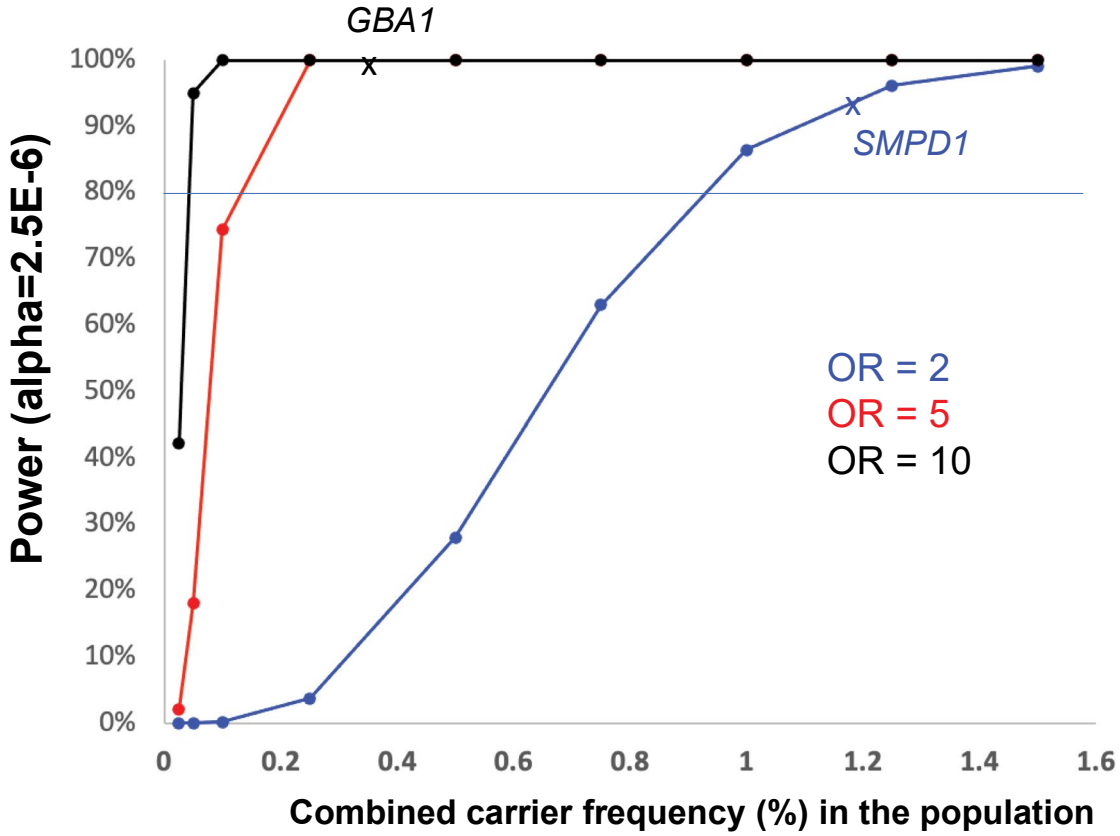
b

Population	Case		Control		No. of variants	OR (95% CI)	pvalue
	Carrier	Non-carrier	Carrier	Non-carrier			
SGMAL	5	2606	43	3701	5	0.17 (0.05-0.42)	
HK	0	70	4	582	4	0 (0-12.81)	
KR	2	1415	1	1039	1	1.47 (0.08-86.68)	
TW	0	200	2	140	1	0 (0-3.77)	
CMH_Discovery						**0.19 (0.08-0.43)**	**9.46e-06**



Extended Data Fig. 7 | Forest plots of PD associations with (a) *HLA-DRB1* qualifying variants and (b) *ATP5B* qualifying variants. CMH without continuity correction used across strata. Data presented as OR and 95% confidence interval (CI).

https://doi.org/10.1038/s43587-024-00760-7



Extended Data Fig. 8 | Statistical power to identify Parkinson's disease-associated genes at exome-wide significance. Statistical power to identify PD-associated genes at exome-wide significance (alpha = 2.5×10^{-6}) is dependent on sample size, the combined carrier frequency in the population and effect size (odds ratio) of the variants in each gene. Based on our current Discovery cohort size of 4,298 PD cases and 5,512 controls, we had >90% power to detect genes like *GBA1* with large effect sizes (OR > 5) and at combined carrier frequencies ~ 0.5% in the population, and also genes like *SMPD1* with smaller effect size (OR ~ 2) but at higher combined carrier frequencies of 1-1.5% in the population (contributed in part by the *SMPD1* p.Pro332Arg variant). Genes with similar variant effect sizes as *SMPD1* may be missed if they are present at lower combined carrier frequencies of 0.9% or less. Similarly, genes with variant effect sizes comparable with *GBA1* will be missed if they are present at combined carrier frequencies of 0.1% or less.

nature portfolio

Corresponding author(s): Jia Nee Foo

Last updated by author(s): Oct 22, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	The codes used in this study can be found on GitHub (https://github.com/fjnlab/PD_exome-seq). Software and software packages used include: R (version 4.0.2) and R packages data.table (version 1.14.0), dplyr (version 1.0.5), tidyr (version 1.1.2), ggplot2 (version 3.3.6), forestplot (version 2.0.1), stats (mantelhaen.test, version 4.3.1), SKAT (version 2.2.5); BWA (version 0.7.17); GATK (version 3.7); Picard (version 2.21.7); VCFtools (version 0.1.16); BCFtools (version 1.9); PLINK (version 1.9); Fratools PERL package (version 1, https://github.com/atks/fratools); Ensembl Variant Effect Predictor (VEP) software (version 104); ANNOVAR (version dated 2020-06-07); VEP plugin LOFTEE (version 1.0.3); PolyPhen-2 (version 2.2.3r406); dbNSFP (version 3.5a); Seurat (version (version 4.3.0.1); scDblFinder (version 1.15.2); clustree (version 0.5.0); PyMol (version 2.5.2).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All unrestricted data are available from the corresponding author upon request. Summary statistics can be found in Supplementary Table 22. Western blots are provided as Supplementary Data. Acid sphingomyelinase enzymatic activity assay absorbance readings are provided as Source data. Individual level genotype data are not publicly available because of consent and privacy issues but will be made available upon reasonable request to J.N.F. (jianee.foo@ntu.edu.sg, response timeframe of 1 month) and following approval by the relevant institutional review board. Midbrain single nuclei RNA-seq datasets used are available from GEO as detailed in Supplementary Table 5. Protein crystallography of ASM tertiary structure is available from Protein Database Bank (ID: 5I85).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Sex of discovery samples were genetically inferred from exome sequencing data (sequencing coverage of X and Y chromosomes) and compared against patient records to ensure concordance.
Reporting on race, ethnicity, or other socially relevant groupings	All discovery samples were recruited from countries in East Asia and the majority self-reported as Chinese or South Korean ethnicity/ancestry. Principal components analysis was carried out using genotypes from sequencing data to remove genetic outliers from admixed or minority ancestry groups for the final analysis (Supplementary Figure 2). Samples from the replication cohorts were from published datasets and specified as samples of Chinese or European ancestry by the authors.
Population characteristics	Participants in this study were from resident East Asian populations in Singapore, Malaysia, Hong Kong, Taiwan and South Korea. All samples in the final analysis was confirmed to be of East Asian (Chinese/South Korean) ancestry by principal components analysis using genotypes derived from exome sequencing data. Subsequent analysis was adjusted on principal components (PC1-PC3) , per sample variant counts and average per sample coverage as covariates. Population characteristics of 4,298 PD cases are as follows: sampling age: 66.6±9.8 years, onset age=61.1±10.9 years, 53.44% males. Population characteristics of 5,512 neurologically normal controls are as follows: 65.2±9.4 years, 46.88% males. Description for each strata can be found in Supplementary Table 2.
Recruitment	Patients with PD together with ancestry- and geographically matched controls were recruited by 6 independent study groups from 5 regions across East Asia (Singapore (SG), Malaysia (MAL), Hong Kong (HK), South Korea (KR) and Taiwan (TW)). Recruitment sites are detailed in Supplementary Table 1.
Ethics oversight	This study was approved by the ethics committees or institutional review boards of the respective institutions (SingHealth Centralized Institutional Review Board CIRB 2002/008/A and 2019/2334 and Nanyang Technological University Institutional Review Board IRB-2016-08-011). Research with post-mortem human brain tissues was approved by the Nanyang Technological University Institutional Review Board (IRB-2016-12-029 and IRB-2018-09-052) and National Research Ethics Committee (08/MRE09/31 and NHS REC No 18/SW/0029).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The discovery cohort consists of 4,298 cases and 5,512 controls from five regions across East Asia (Singapore, Malaysia, Hong Kong, South Korea and Taiwan). We had >90% power to detect genes with large effect sizes (OR>5) and at combined carrier frequencies ~0.5% in the population, and also genes with smaller effect size (OR ~2) but at higher combined carrier frequencies of 1-1.5% in the population. Refer to Extended Fig. 8 for statistical power calculations.
Data exclusions	4,915 cases and 5,971 controls were recruited for this study. 617 cases and 459 controls did not pass sample quality control and were excluded from the final analysis. 4,298 cases and 5,512 controls remained after sample quality control. Details of sample numbers that failed each step of sample quality control is listed in Supplementary Table 3.
Replication	Published sequencing data from a total of 5,585 PD patients and 5,642 controls from China, United States, and the International PD Genomics

Replication	Consortium (samples ascertained from the United States and from Europe) were evaluated for replication of significant findings from the discovery exome sequencing analysis. See Supplementary Table 15 for details of replication cohorts. The association of rare SMPD1 variants with less than 43.58% of acid sphingomyelinase activity level was successfully replicated in the replication cohort.
Randomization	Sample allocation into case and control groups was not random and based upon Parkinson's disease diagnosis. A diagnosis of Parkinson's disease was reached using the United Kingdom Parkinson's Society Brain Bank Criteria. Conversely, the controls were healthy subjects that do not have any overt neurological diseases.
Blinding	Variant calling/identification and annotation was carried out blinded to affection status of samples. Functional testing of SMPD1 variants was also carried out blinded to affection status of the sample in which they were identified. Blinding was not possible during the association analysis as samples had to be classified into case or controls for the analysis process.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Primary Antibodies: SMPD1 antibody (R&D Systems, Catalog number: MAB5348, Clone name: 563418, Lot number: CEIM0121101) GBA1 antibody (Santa Cruz, Catalog number: sc-166407, Clone name: B-6, Lot number: I0122) GAPDH antibody (Bio-rad, Catalog number: VMA00046, Clone name: NA, Lot number: 151149) TH antibody (Abcam, Catalog number: ab137869, Clone name: EP1532Y, Lot number: GR3370594-6) HuC/D antibody (Life Technologies, Catalog number: A-21271, Clone name: 16A11, Lot number: 1963099) Secondary Antibodies: Goat anti-mouse Alexa Fluor Plus 488 (ThermoFisher Scientific, Catalog number:A32723, Lot number: YH376193)
Validation	All primary antibodies were used according to manufacturer's recommended dilutions for immunohistochemistry or Western blotting, and validated for compatibility with Human by the manufacturer or previous publications as documented by the manufacturer. Primary Antibody / Published Species / Application / Citations SMPD1 antibody (MAB5348) (https://www.rndsystems.com/products/human-smpd1-antibody-563418_mab5348) / Human / Immunoprecipitation / Cited in 1 publication for western blotting GBA1 antibody (sc-166407) (https://www.scbt.com/p/beta-glucosidase-antibody-b-6?srsltid=AfmBOoqL9Y7boaemTSx-9WWa2LGJakluiCWl7JdbFrbgtmDuq2LNrEFJ) / Mouse, Rat and Human / WB, IP, IF, IHC(P), ELISA / Cited in 7 publications GAPDH antibody (VMA00046) (https://www.scbt.com/p/gapdh-antibody-0411?gad_source=1&gclid=Cj0KCQjwsJO4BhDoARIsADDv4vAnhtOrlT_tVd0fFbOCPjhF3mhkv2OcTJtk1fJJrI6TAbXc65UKKY0aAo84EALw_wcB) / Human / WB, IP, IF and IHC(P) / Cited in 3,228 publications TH antibody (ab137869) (https://www.abcam.com/en-us/products/primary-antibodies/tyrosine-hydroxylase-antibody-ep1532y-neuronal-marker-ab137869) / Mouse, Rat, Human / IHC-P, WB, ICC/IF / Cited in 69 publications HuC/D antibody (A-21271) (https://www.thermofisher.com/antibody/product/HuC-HuD-Antibody-clone-16A11-Monoclonal/A-21271) / Avian, Chicken, Human, Zebrafish / WB, IHCC, IF, Flow, ELISA / Cited in 15 publications for western blotting and 323 publications for IHC

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	293T from ATCC, catalog number ATCC CRL-3216
Authentication	STR profiling conducted by ATCC

Mycoplasma contamination	All cell lines tested negative for mycoplasma contamination
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were used in this study

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*