Mitochondrial haplogroups and cognitive progression in Parkinson's disease

Supplementary material

List of Acknowledgements

Harvard Biomarkers Study (including HBS and HBS2 parts). Co-Directors: Brigham and Women's Hospital: Clemens R. Scherzer, Massachusetts General Hospital: Bradley T. Hyman; Investigators and Study Coordinators: Brigham and Women's Hospital: Yuliya Kuras, Karbi Choudhury, Nada Laroussi, Daly Franco, Michael T. Hayes, Nutan Sharma, Vikram Khurana, Claudio Melo De Gusmao, Chizoba C. Umeh, Reisa Sperling; Massachusetts General Hospital: John H. Growdon, Michael A. Schwarzschild, Albert Y. Hung, Aleksandar Videnovic, Alice W. Flaherty, Deborah Blacker, Anne-Marie Wills, Steven E. Arnold, Ann L. Hunt, Nicte I. Mejia, Anand Viswanathan, Stephen N. Gomperts, Mark W. Albers, Maria Allora-Palli, David Hsu, Alexandra Kimball, Scott McGinnis, John Becker, Randy Buckner, Thomas Byrne, Maura Copeland, Bradford Dickerson, Matthew Frosch, Theresa Gomez-Isla, Steven Greenberg, Julius Hedden, Elizabeth Hedley-Whyte, Keith Johnson, Raymond Kelleher, Aaron Koenig, Maria Marquis-Sayagues, Gad Marshall, Sergi Martinez-Ramirez, Donald McLaren, Olivia Okereke, Elena Ratti, Christopher William, Koene Van Dij, Shuko Takeda, Anat Stemmer-Rachaminov, Jessica Kloppenburg, Catherine Munro, Rachel Schmid, Sarah Wigman, Sara Wlodarcsyk; Data Coordination: Brigham and Women's Hospital: Thomas Yi; Biobank Management Staff: Brigham and Women's Hospital: Idil Tuncali. We thank all study participants and their families for their invaluable contributions. HBS is made possible by generous support from the Harvard NeuroDiscovery Center, with additional contributions from the Michael J Fox Foundation, NINDS U01NS082157, U01NS100603, and the Massachusetts Alzheimer's Disease Research Center NIA P50AG005134.

DIGPD: Steering committee: Jean-Christophe Corvol (Pitié-Salpêtrière Hospital, Paris, principal investigator of DIGPD), Alexis Elbaz (CESP, Villejuif, member of the steering committee), Marie Vidailhet (Pitié-Salpêtrière Hospital, Paris, member of the steering committee), Alexis Brice (Pitié-Salpêtrière Hospital, Paris, member of the steering committee and PI for genetic analysis) ; Statistical analyses: Alexis Elbaz (CESP, Villejuif, PI for statistical analyses), Fanny Artaud (CESP, Villejuif, statistician); Principal investigators for sites (alphabetical order): Frédéric Bourdain (CH Foch, Suresnes, PI for site), Jean-Philippe Brandel (Fondation Rothschild, Paris, PI for site), Jean-Christophe Corvol (Pitié-Salpêtrière Hospital, Paris, PI for site), Pascal Derkinderen (CHU Nantes, PI for site), Franck Durif (CHU Clermont-Ferrand, PI for site), Richard Levy (CHU Saint-Antoine, Paris, PI for site), Fernando Pico (CH Versailles, PI for site), Olivier Rascol (CHU Toulouse, PI for site); Co-investigators (alphabtical order): Anne-Marie

Bonnet (Pitié-Salpêtrière Hospital, Paris, site investigator), Cecilia Bonnet (Pitié-Salpêtrière Hospital, Paris, site investigator), Christine Brefel-Courbon (CHU Toulouse, site investigator), Florence Cormier-Dequaire (Pitié-Salpêtrière Hospital, Paris, site investigator), Bertrand Degos (Pitié-Salpêtrière Hospital, site investigator), Bérangère Debilly (CHU Clermont-Ferrand, site investigator), Alexis Elbaz (Pitié-Salpêtrière Hospital, Paris, site investigator), Monique Galitsky (CHU de Toulouse, site investigator), David Grabli (Pitié-Salpêtrière Hospital, Paris, site investigator), Andreas Hartmann (Pitié-Salpêtrière Hospital, Paris, site investigator), Stephan Klebe (Pitié-Salpêtrière Hospital, Paris, site investigator), Julia Kraemmer (Pitié-Salpêtrière Hospital, site investigator), Lucette Lacomblez (Pitié-Salpêtrière Hospital, Paris, site investigator), Sara Leder (Pitié-Salpêtrière Hospital, Paris, site investigator), Graziella Mangone (Pitié-Salpêtrière Hospital, Paris, site investigator), Louise-Laure Mariani (Pitié-Salpêtrière Hospital, Paris, site investigator), Ana-Raquel Marques (CHU Clermont Ferrand, site investigator), Valérie Mesnage (CHU Saint Antoine, Paris, site investigator), Julia Muellner (Pitié-Salpêtrière Hospital, Paris, site investigator), Fabienne Ory-Magne (CHU Toulouse, site investigator), Violaine Planté-Bordeneuve (Henri Mondor Hospital, Créteil, site investigator), Emmanuel Roze (Pitié-Salpêtrière Hospital, Paris, site investigator), Melissa Tir (CH Versailles, site investigator), Marie Vidailhet (Pitié-Salpêtrière Hospital, Paris, site investigator), Hana You (Pitié-Salpêtrière Hospital, Paris, site investigator); Neuropsychologists: Eve Benchetrit (Pitié-Salpêtrière Hospital, Paris, neuropsychologist), Julie Socha (Pitié-Salpêtrière Hospital, Paris, neuropsychologist), Fanny Pineau (Pitié-Salpêtrière Hospital, Paris, neuropsychologist), Tiphaine Vidal (CHU Clermont-Ferrand, neuropsychologist), Elsa Pomies (CHU de Toulouse, neuropsychologist), Virginie Bayet (CHU de Toulouse, neuropsychologist); Genetic core: Alexis Brice (Pitié-Salpêtrière Hospital, Paris, PI for genetic studies), Suzanne Lesage (INSERM, ICM, Paris, genetic analyses), Khadija Tahiri (INSERM, ICM, Paris, lab technician) Hélène Bertrand (INSERM, ICM, Paris, lab technician), Graziella Mangone (Pitié-Salpêtrière Hospital, Paris, genetic analyses); Sponsor activities and clinical research assistants: Alain Mallet (Pitié-Salpêtrière Hospital, Paris, sponsor representative), Coralie Villeret (Hôpital Saint Louis, Paris, Project manager), Merry Mazmanian (Pitié-Salpêtrière Hospital, Paris, project manager), Hakima Manseur (Pitié-Salpêtrière Hospital, Paris, clinical research assistant), Mostafa Hajji (Pitié-Salpêtrière Hospital, Paris, data manager), Benjamin Le Toullec (Pitié-Salpêtrière Hospital, Paris, clinical research assistant), Vanessa Brochard (Pitié-Salpêtrière Hospital, Paris, project manager), Monica Roy (CHU de Nantes, clinical researh assistant), Isabelle Rieu (CHU Clermont-Ferrand, clinical research assistant), Stéphane Bernard (CHU Clermont-Ferrand, clinical research assistant), Antoine Faurie-Grepon (CHU de Toulouse, clnical research assistant). The study was sponsored

by the Assistance Publique Hôpitaux de Paris, and was funded by a grant from the Ministry of Health (PHRC AOR0810).

PreCEPT/PostCEPT Study: PreCEPT/PostCEPT Steering Committee: University of Rochester: David Oakes, Ira Shoulson; University of Toronto: Anthony E. Lang; Parlinson's Institute: Caroline Tanner; Institute for Neurodegenerative Disorders: Kenneth Marek; Voyager Therapeutics: Bernard Ravina; Brigham and Women's Hospital: Clemens Scherzer, University of Ottawa: Michael Schlossmacher, Avid Radiopharmaceuticals: Andrew Siderowf, We thank the Parkinson Study Group (PSG) PreCEPT/PostCEPT investigators for the acquisition of high-quality clinical data, careful follow up of study subjects and collection of blood samples.

DATATOP: We thank the investigators of the Parkinson Study Group (PSG) DATATOP for the acquisition of high-quality clinical data, careful follow-up of study subjects, and DNA collection in the DATATOP cohort.

PICNICS: Investigators:Roger Barker, Caroline Williams-Gray, David P Breen, Gemma Cummins, Jonathan Evans, Sophie Winder-Rhodes, Ruwani Wijeyekoon. The PICNICS study was sponsored by the University of Cambridge/Cambridge University Hospitals NHS Trust UK and received funding from the Cure Parkinson's Trust, the Van Geest Foundation, Parkinson's UK, and the NIHR Cambridge Biomedical Research Centre.

CamPaIGN: Investigators: Roger Barker, Tom Foltynie, Caroline Williams-Gray, Trevor Robbins, Carol Brayne, Sarah Mason, Sophie Winder-Rhodes, Ruwani Wijeyekoon. The CamPaIGN study was sponsored by the University of Cambridge/Cambridge University Hospitals NHS Trust UK and has received funding from the Wellcome Trust, the Medical Research Council, the Patrick Berthoud Trust, and the NIHR Cambridge Biomedical Research Centre.

PROPARK/PROPARK-C: The PROPARK study was headed by Jacobus J. van Hilten and Johan Marinus.

PDBP: Data and biospecimens used in preparation of this manuscript were obtained from the Parkinson's Disease Biomarkers Program (PDBP) Consortium, supported by the National Institute of Neurological Disorders and Stroke at the National Institutes of Health. Investigators include: Roger Albin, Roy Alcalay, Alberto Ascherio, Thomas Beach, Sarah Berman, Bradley

Boeve, F. DuBois Bowman, Shu Chen, Alice Chen-Plotkin, William Dauer, Ted Dawson, Paula Desplats, Richard Dewey, Ray Dorsey, Jori Fleisher, Kirk Frey, Douglas Galasko, James Galvin, Dwight German, Lawrence Honig, Xuemei Huang, David Irwin, Kejal Kantarci, Anumantha Kanthasamy, Daniel Kaufer, James Leverenz, Carol Lippa, Irene Litvan, Oscar Lopez, Jian Ma, Lara Mangravite, Karen Marder, Laurie Orzelius, Steven Gunzler, Vladislav Petyuk, Judith Potashkin, Liana Rosenthal, Rachel Saunders-Pullman, Clemens Scherzer, Michael Schwarzschild, Tanya Simuni, Andrew Singleton, David Standaert, Debby Tsuang, David Vaillancourt, David Walt, Andrew West, Cyrus Zabetian, Jing Zhang, and Wenquan Zou. The PDBP Investigators have not participated in reviewing the data analysis or content of the manuscript.

PPMI: Data used in the preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database (www.ppmi-info.org/data). For up-to-date information on the study, visit www.ppmi-info.org. PPMI, a public-private partnership, is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including Abbvie, AcureX, Allergan, Amathus, Asap, Avid, Bial Biotech, Biogen Idec, BioLegend, Bristol-Myers Squibb, Calico, Celgene, DACAPO Brainscience, DENALI, 4D pharma plc, EDMOND J. SAFRA, GE Healthcare, Genentech, GlaxoSmithKline, GOLUB CAPITAL, Handl Therapeutics, Insitro, Janssen Neuroscience, Lilly, Lundbeck, Merck, Meso Scale Discovery, Neurocrine, Pfizer, Piramal, Prevail Therapeutics, Roche, SANOFI GENZYME, SERVIER, Takeda, TEVA, UCB, Verily and Voyager Therapeutics.

Arizona Study of Aging/Brain and Body Donation Program: National Institute of Neurological Disorders and Stroke, U24 NS072026 National Brain and Tissue Resource for Parkinson's Disease and Related Disorders; National Institute on Aging, P30 AG19610 Arizona Alzheimer's Disease Core Center; Arizona Department of Health Services, Arizona Alzheimer's Consortium; Arizona Biomedical Research Commission, Arizona Parkinson's Disease Consortium; Michael J. Fox Foundation for Parkinson's Research.

NET-PD LS1: We would like to thank the patients and families who participate in the NET-PD LS1 study. The following additional NINDS grants supported the Net-PD LS-1 study: U01NS043127, U01NS043128, and U10NS44415-44555 from the National Institute of Neurologic Disorders and Stroke.

Tartu/Perron: We would like to thank all patients and families who participated in the study. Support by the following grants and funders: institutional research grants PRG957 and IUT20–46 of the Estonian Research Council, H2020 ERA-chair grant (agreement 668989, project Transgeno), MSWA, The Michael J. Fox Foundation, Shake It Up Australia and The Perron Institute.

ParkWest: Principal investigators: Guido Alves (Norwegian Centre for Movement Disorders, Stavanger University Hospital), Ole-Bjørn Tysnes (Haukeland University Hospital). Investigators and study coordinators: Karen Herlofson, Solgunn Ongre, Siri Bruun (Sørlandet Hospital Arendal); Ineke HogenEsch, Marianne Kjerandsen, Liv Kari Håland (Haugesund Hospital); Wenche Telstad, Aliaksei Labusau, Jane Kastet (Førde Hospital); Bernd Müller, Geir Olve Skeie, Charalampos Tzoulis (Haukeland University Hospital); Kenn Freddy Pedersen, Michaela Dreetz Gjerstad, Elin Bjelland Forsaa, Jodi Maple-Grødem, Johannes Lange, Veslemøy Hamre Frantzen, Anita Laugaland, Karen Simonsen, Ingvild Dalen (Stavanger University Hospital). The ParkWest study has received funding from the Western Norway Regional Health Authority (grant number 911218), and the Norwegian Parkinson's Disease Association, and the Research Council of Norway (grant number 287842).

PIB funding: NIH grants: NS075321 and NS097437, the American Parkinson Disease Association (APDA), the Greater St. Louis Chapter of the APDA, the Barnes Jewish Hospital Foundation (Elliot Stein Family Fund, Parkinson disease research fund).

Supplementary Methods

Study participants

4,491 patients with PD (with available genotyping data and quality control) were longitudinally assessed with 3,3406 study visits in 15 cohorts from North America and Europe between 1986 and 2017: Harvard Biomarkers Study (HBS)[1], Neuroprotection Exploratory Trials in PD- Long term Study-1 (NET-PD LS1)[2], Drug Interaction with Genes in PD (DIGPD)[3] , PROfiling PARKinson's disease (PROPARK) study[4], PROPARK-Cross sectional cohort; Cambridgeshire Parkinson's Incidence from GP to Neurologist (CamPaIGN)[5-7]; Parkinsonism: Incidence, Cognition and Non-motor heterogeneity in Cambridgeshire (PICNICS)[8]; Parkinson's Disease Biomarkers Program (PDBP)[9]; Banner Health study(Arizona Study of Aging/Brain and Body Donation Program)[10]; ParkWest[11] and PIB[12]; Deprenyl and Tocopherol Antioxidative Therapy of Parkinsonism (DATATOP)[13]; Parkinson Research Examination of CEP-1347 Trial/A Longitudinal Follow-up of the PRECEPT Study Cohort (PreCEPT/PostCEPT)[14] and Tartu[15], Parkinson's Progression Markers Initiative (PPMI)[16] . For PPMI, approval was obtained to download and analyze the publicly accessible WGS and clinical data. 13 cohorts enrolled patients with a diagnosis of PD established according to modified UK PD Society Brain Bank diagnostic criteria as previously reported[1-5,8,9,11,12,14-18]. In DATATOP, the eligibility criteria required a clinical diagnosis of early, idiopathic PD (HY stages 1 or 2) with patients not on anti-parkinsonian medications[17]. Banner Health study: all subjects have come to autopsy and have had full neuropathological examinations with diagnosis[10]. Diagnostic certainty was increased by confirming the clinical diagnosis of PD during longitudinal follow-up visits[19] in all cohorts.

Serial Mini Mental State Exam (MMSE) scores[20] were longitudinally collected in 10 cohorts. Montreal Cognitive Assessment (MoCA)[21] scores were collected in the PDBP, PPMI study and converted to MMSE scores according to a published formula[22]. SCOPA-COG were collected in PROPARK, PROPARK-C and NET-PD LS1 cohort and converted to MMSE scores.

Polymorphism identification and haplogroup classification

The genotyping data of the 4,491 subjects with Parkinson's disease were reported in Ref.[23]. Briefly, the samples (excluded PPMI with WGS) were genotyped with Illumina Multi-Ethnic Genotyping Array (MEGA, Illumina), which includes 810 SNP markers in mtDNA after quality control as described in Ref.[23] 810 mtDNA variants were converted from "plink" format to "vcf" format according to the rCRS reference alleles. We removed 25 mismatched SNPs and InDel SNPs, 11 duplicated SNP probes on the array, and 11 variants with highly discordant MAF (> 5%) compared to Phase 1 and 3 of the 1000 Genomes Project[24] mitochondrial variants (n = 503 European) as called by the MToolBox pipeline[25]. The remained 763 SNPs were used to predict mitochondrial haplogroups using Haplogrep2.0[26] with default parameter using rCRS reference. Haplotype quality-control was performed according to the Haplogrep2 instruction and 44 subjects with quality score < 0.8 were excluded (**Supplementary Fig. 1**). We next simplified the sub-haplogroups (455 sub-haplogroups) to the 34 haplogroups (**Supplementary Table 1**) according to the mtDNA tree http://www.phylotree.org/tree/index.htm. 4,447 subjects were successfully assigned haplogroup, and 24 of these patients had no clinical records of note so were removed from the analysis. We further removed the haplogroups with less than 100 subjects (**Supplementary Table 1**, total 359 subjects), and the remaining 4,064 subjects with 30,515 study visits were used for haplogroup analysis (including H, HV* (excluding H, V), I, J, K, T, U#

(excluding K) haplogroups). It should be note that in our work, U# denotes all U haplogroups (U1, U2, U3, U4, U5, U6, U7…) but not haplogroup K as in Ref[27]. Out of 763 SNPs, 102 SNPs with allele frequency > 1% were used for single SNP Cox regression analysis. Notably, common mtDNA haplogroup or mtDNA variants are often ancient and are usually homoplasmic[28]. We did not analyze heteroplasmic mtDNA mutations in this study.

Statistical analysis

The Cox proportional hazards statistic was used to estimate the influence of different mitochondrial haplogroups on time (years from onset of PD) to reaching the endpoint of motor disability with postural instability (Hoehn and Yahr stage HY 3) or global cognitive impairment (GCI) as indicated by a MMSE ≤ 25 according to the recommendation the International Parkinson and Movement Disorder Society (MDS) Task Force[29] as in Ref.[30]. For HY analysis, age at onset of PD, sex and *GBA* carrier status were included as covariates. In the GCI analysis, age at onset of PD, sex, years of education, and polygenic hazard score (PHS including *GBA* carrier status, *APOE* ε4 allele haplotype and three novel progression variants rs182987047, rs138073281 and rs8050111 from Ref.[23]) were included as covariates in the Cox analyses. A "cohort" term was included as a random effect (a random effects Cox model is often termed a "frailty" model). 29,115 (95.4 %) of the visits from 4,088 patients with PD occurred within 12 years of longitudinal follow-up from disease onset with a median follow-up time of 6.7 years (inter-quartile range, 4.2 years), thus we focused our survival analyses on the 12-year time frame from disease onset. Patients were left-censored and those with missing or non-quality clinical data were excluded. Cox proportional hazards analyses were performed using the coxph function in the Survival package (v2.38-1)[31] and the "breslow" method was used for handling observations that have tied

survival times in the analysis and *P* values less than or equal to 0.05 were considered as indicative of haplogroup significance.

For single polymorphism variants analysis, we used a similar Cox proportional hazards regression model (same co-variants as mentioned above) to investigate each SNP effect on motor and cognitive impairment. Bonferroni correction was performed using p.adjust function with "bonferroni" in R.

Generalized longitudinal mixed fixed and random effects analysis (LMM)[32] of cognitive decline was performed with serial Mini Mental State Exam (MMSE) scores longitudinally assessed at varying times (enrollment visit and multiple longitudinal follow-up visits) in the combined data set. Two cohorts (PROPARK-C and Tartu) were excluded from the LMM because no longitudinal MMSE scores were available. The MMSE score was the dependent variable and the primary predictors were mitochondrial haplogroup status, time in the study (years), and their interaction. An intercept term and linear rate of change across time per subject were the random terms (permitted to be correlated). Subject level fixed covariates were age at baseline, sex, years of education, duration of PD illness at baseline, as well as PHS score. A study term was included as a random effect. This analysis was performed using the lme4 package (v1.1-23). All analyses were conducted in the R statistical environment version 4.0.2.

Comparison of models

The original multivariable Cox model from a previous study[33] included age at Parkinson's disease onset, years of education, sex, MMSE at enrolment, MDS-UPDRS III score at enrolment, depression at enrolment and *GBA* carrier status, and a cohort term was included as a random effect (using a frailty Cox model). 2,629 patients in the original nine longitudinal cohorts with available

mitochondrial variants, and 2,376 patients (253 left censored patients were removed) with 22,617 visits within 12 years of longitudinal follow-up from disease onset were used for comparison of different Cox regression genetic models (*GBA* carrier, *APOE* ε4, m.2706A>G, m.14766C>T), adjusting by age at Parkinson's disease onset, years of education, sex, MMSE at enrolment, MDS-UPDRS III score at enrolment, depression at enrolment, and a cohort term was included as a random effect.

Combination analysis of two genetic risk (*GBA* carrier and m.2706A>G variant/*APOE* ε4 and m.2706A>G variant) was performed using 2,376 patients from the Cox regression model, adjusting by the same six clinical predictors as mentioned bove, and a cohort term was included as a random effect.

Supplementary Table 1 Classification of mitochondrial haplogroups in patients with PD across the 15 cohorts

Haplogroups	Number	Sub-haplogroups of H	Number
H	1,829		
U#	666	H1	599
T	440	H2	599
J	427	H3	155
K	393	H5	146
HV*	218	H6	94
I	115	H4	75
W	93	H13	47
X	75	H46	20
N1	65	H15	11
V	42	H26	11
R0[&]	17	H7	10
L2	11	H41	8
M1	9	H14	6
D	7	H79	6
A	5	H85	6
C	5	H28	5
L1	4	H44	5
R1	4	H24	4
L3e	3	H100	3
B	2	H22	3
L3b	2	H81	3
M9	2	H56	2
N2	2	H94	2
N3	2	H17	1
G	1	H30	1
L0	1	H33	1
M7	1	H34	1
M30	1	H42	1
M33	1	H49	1
M49	1	H50	1
N9	1	H73	1
Y	1	H77	1
Z	1		

Haplogroups according to the mtDNA tree http://www.phylotree.org/tree/index.htm. 4,447 subjects were successfully assigned haplogroup. HV*: not including H, HV; U[#]: including all U haplogroups (U1, U2, U3, U4, U5, U6, U7…) but not haplogroup K as in Ref[27]; R0[&]: not including HV, H, V.

Supplementary Table 2 Clinical characteristics of patients with PD at enrollment with different macro-haplogroups across the 15 cohorts

***n* = 4,423**	**Marco-haplogroup**								***P*[&]**
	H	**HV***	**I**	**J**	**K**	**T**	**U[#]**	**Others**	
Number of men (*n*, %)	1,819 (63.2)	217 (65.4)	115 (57.4)	427 (61.8)	391 (65.7)	435 (62.1)	660 (63.8)	359 (62.1)	0.76
Age at onset, mean (SD), years	60.6 (10.6)	61.4 (10.1)	61.0 (10.8)	61.6 (10.9)	61.1 (9.6)	60.9 (10.5)	60.7 (11.0)	61.3 (10.1)	0.45
Age at enrollment, mean (SD), years	64.2 (10.2)	64.8 (9.8)	64.5 (10.5)	65.1 (10.4)	64.2 (9.3)	64.4 (10.1)	64.3 (10.5)	64.8 (10.3)	0.62
Years of education, mean (SD), years	14.2 (3.8)	14.6 (3.7)	14.1 (4.1)	14.4 (3.9)	14.6 (3.8)	14.1 (3.9)	14.2 (3.6)	14.7 (3.6)	0.08
Study years, mean (range), years	3.8 (0-19.9)	3.7 (0-9.3)	3.9 (0-8.3)	3.6 (0-13.1)	3.8 (0-14.5)	3.6 (0-13.5)	3.6 (0-12.6)	3.5 (0-12.3)	0.26
Hoehn and Yahr, mean (SD)	1.9 (0.8)	1.9 (0.7)	1.8 (0.6)	1.9 (0.8)	1.9 (0.7)	2.0 (0.7)	2.0 (0.7)	1.9 (0.7)	0.28
MDS-UPDRS III, mean (SD)	28.4 (14.2)	27.2 (13.5)	27.4 (13.7)	27.9 (14.5)	26.9 (13.0)	28.4 (13.7)	28.4 (14.0)	27.9 (14.9)	0.72
MMSE, mean(SD)	28.2 (2.2)	28.2 (1.9)	28.5 (1.5)	28.1 (2.3)	28.3 (2.0)	28.2 (2.1)	28.2 (2.2)	28.3 (2.4)	0.47
LED, mean(SD)	436.6 (439.9)	402.3 (470.0)	428.6 (458.2)	415.2 (428.0)	373.6 (398.2)	399.9 (416.8)	433.2 (446.0)	418.7 (447.7)	0.34

24 subjects have no available clinic data, the table showed clinical characteristics of 4,423 patients with PD.

[&] Fisher exact test was used for the number of men in each group. Group comparisons were performed using Kruskal-Wallis test for age at onset, age at enrollment, years of education, study years, HY, MDS-UPDRS III, MMSE, LED.

HV*: The sub-haplogroups of haplogroup HV, not including haplogroup H, V.

U[#]: The sub-haplogroups of haplogroup U, not including haplogroup K.

Supplementary Table 3 The percentage (%) of different mitochondrial haplogroups in European population from literatures

Haplogroup	Latvia[34] n=299	Spain[35] n=312	Portugal[36] n=241	France[37] n=210	Norway[34] n=397	Czech[38] n=300	Germany[34] n=333	Iceland[27] n=467	Italy[39] n=124
H	44.5	42.3	40.7	41.9	45.1	40.7	47.7	47.6	41.1
HV*	2.3	NA	NA	NA	0.3	2.7	0.6	NA	1.6
I	4.3	1.6	0.8	2.9	2.3	2.0	1.8	4.7	NA
J	6.4	6.7	6.6	5.2	12.6	8.3	8.4	14.1	4.8
K	2.3	4.8	5.4	11.4	5.0	4.0	7.5	7.7	1.6
T	9.4	8.3	10.8	11.9	9.8	8.0	9.0	10.1	8.1
U[#]	23.1	16.0	17.4	17.6	16.9	21.3	13.5	11.8	30.6

HV*: The sub-haplogroups of haplogroup HV, not including haplogroup H, V;
H: sum of available sub-haplogroups of H;
J: sum of available sub-haplogroups of J;
K: sum of available sub-haplogroups of K;
T: sum of available sub-haplogroups of T;
U[#]: sum of available sub-haplogroups of haplogroup U, but not including haplogroup K.

Supplementary Table 4 Test for residual heterogeneity for each haplogroup compared to haplogroups of H in GCI combined analysis

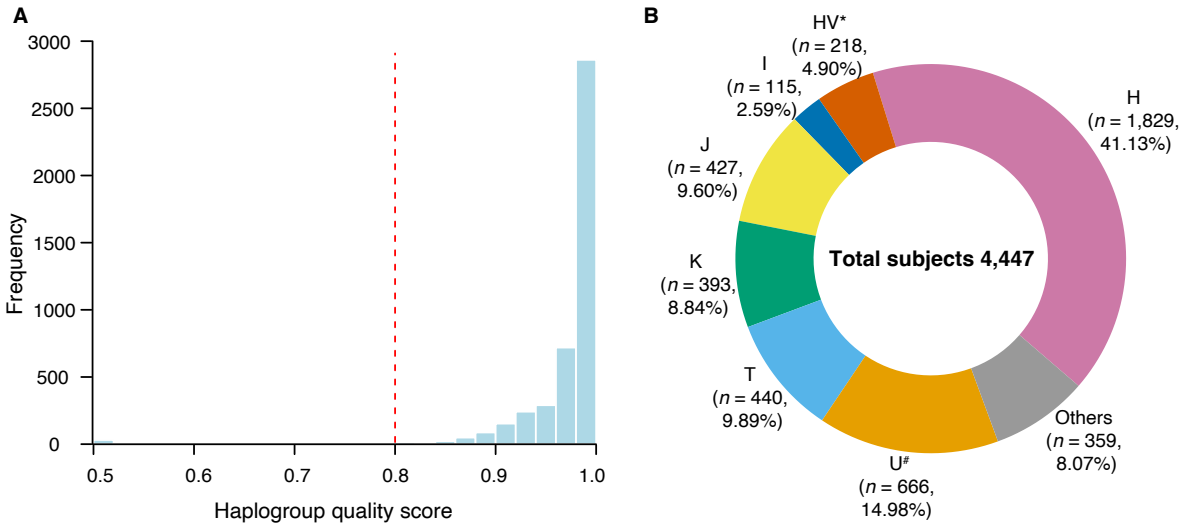
Haplogroups (H as reference)	Heterogeneity Q	*P* value[&]	I^2
HV*	5.32	0.87	0%
I	12.59	0.32	12.61%
J	4.88	0.96	0%
K	7.54	0.82	0%
T	5.05	0.96	0%
U[#]	15.90	0.20	24.54%

HV*: The sub-haplogroups of haplogroup HV, not including haplogroup H, V.
U[#]: The sub-haplogroups of haplogroup U, not including haplogroup K.
[&]The Cochran's Q-test was used to test for residual heterogeneity across studies via R metafor package (version 2.4-0). I^2 index (100%×(Q-df)/Q) was used to quantify the degree of heterogeneity.

Supplementary Table 5 The association of mitochondrial haplogroups in Alzheimer disease

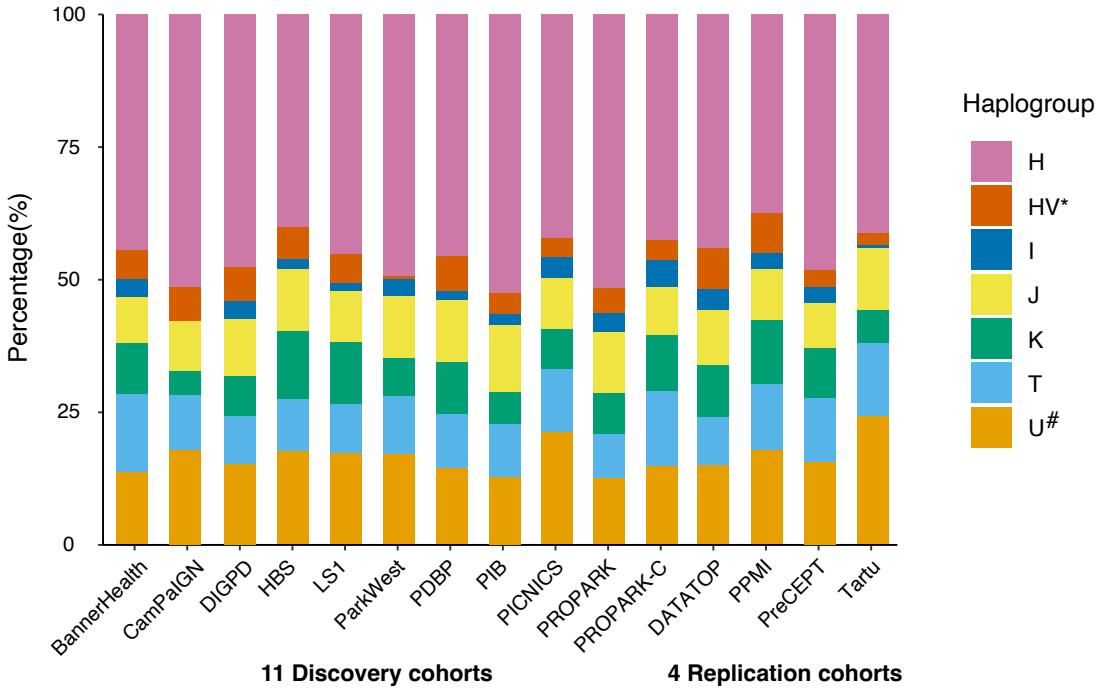
Haplogroup	Years	Effect	Ethnicity	Dataset size (case/control)	Dataset type
H	2009[40]	Risk	Poland	222/252	Whole mitochondrial genomics
	2011[41]	Risk	Caucasian	422/318	Control_region position (16624-576) + 9 coding SNPs
HV	2009[40]	Risk	Poland	222/252	Whole mitochondrial genomics
	2011[41]	Risk	Caucasian	422/318	Control_region position (16624-576) + 9 coding SNPs
K	2001[42]	Protective	Italian	213/389	10 restricted sites
	2011[41]	Protective	Caucasian	422/318	Control_region position (16624-576) + 9 coding SNPs
	2020[43]	Protective	American	309/507	Whole mitochondrial genomics
K1A1B	2013[44]	Risk	Caucasian	154/175	138SNPS
J	2009[40]	Protective in males	Poland	222/252	Whole mitochondrial genomics
	2020[43]	Risk	American	309/507	Whole mitochondrial genomics
T	2011[41]	Protective in females	Caucasian	422/318	Control_region position (16624-576) + 9 coding SNPs
JT	2011[41]	Protective in females	Caucasian	422/318	Control_region position (16624-576) + 9 coding SNPs
U	2001[42]	Protective	Italian	213/389	10 restricted sites
	2004[45]	Risk in males, protective in females	Caucasian	989/328	10 SNPs

Supplementary Figure 1 The classification of haplogroup in patients with PD across 15 cohorts.



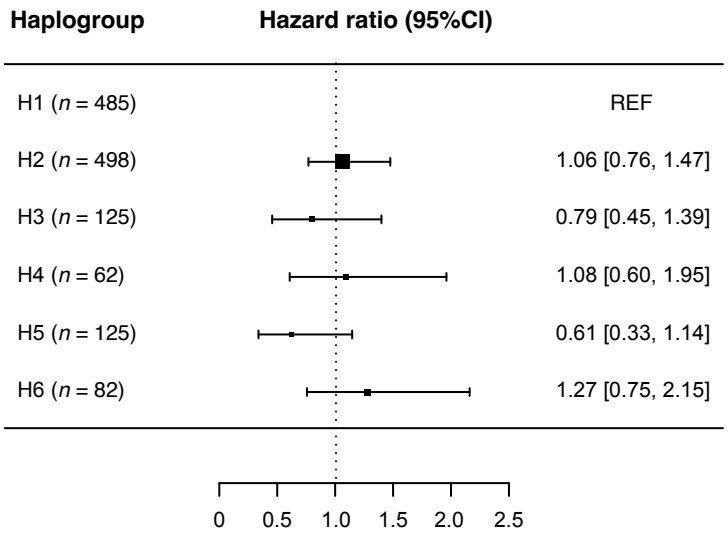
A The haplogroup quality score of 4,491 patients with PD was evaluated from HaploGrep2.0[26] based on Kulczynski measure: (HaplogroupWeight + SampleWeight) × 0.5. The HaploGrep2.0 applied this formula to all haplogroups in Phylotree and returned the overall best hit and the score represented its haplogroup quality. The quality of 0.8 as cutoff was recommended and 4,447 subjects were successfully assigned mitochondrial haplogroup. **B** The donut plot presents the proportion of patients with PD within diverse mitochondrial macro-haplogroups

Supplementary Figure 2 The stacking diagram for distribution of seven macro-haplogroups in patients with PD across 15 cohorts



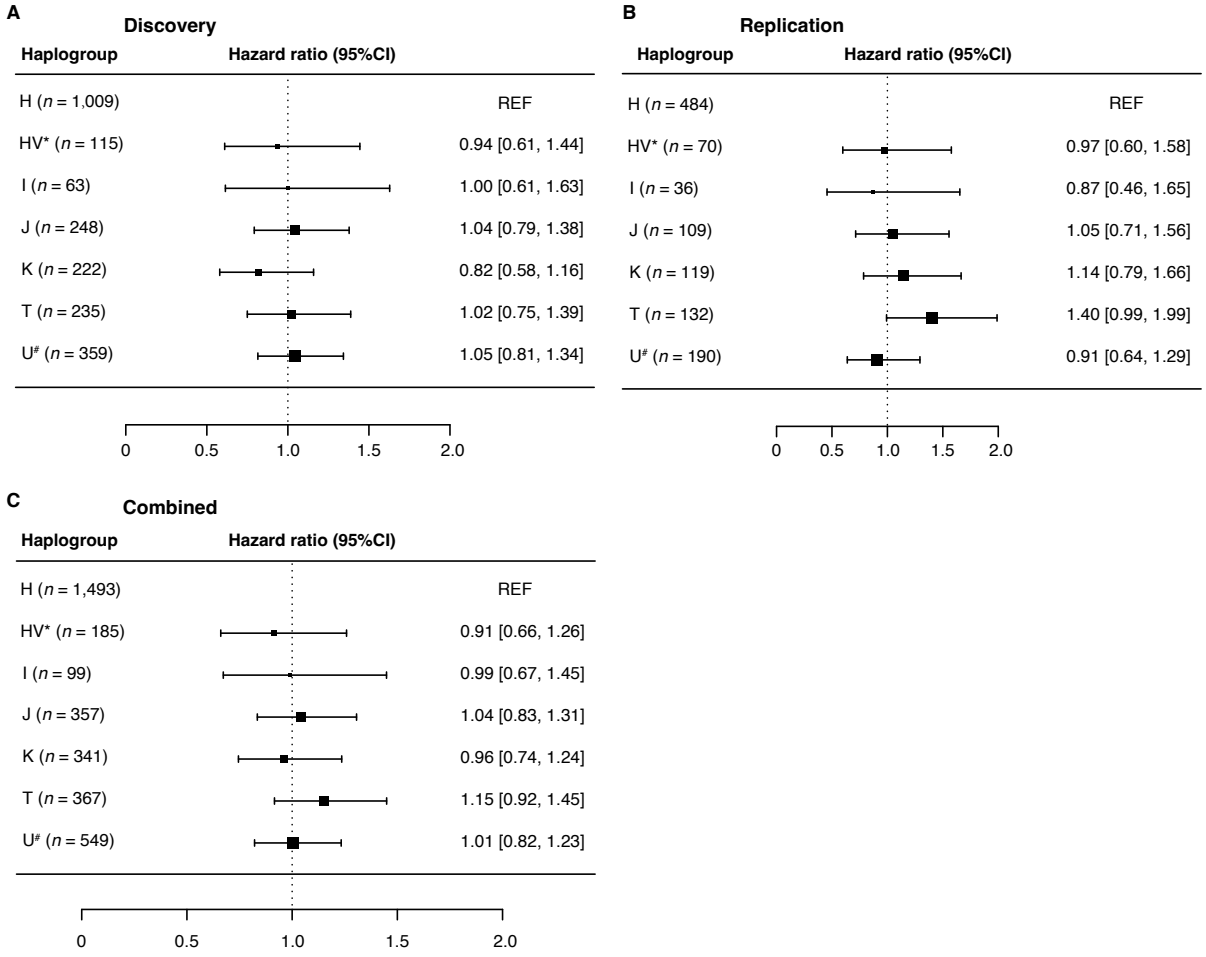
Each vertical bar corresponds to one cohort and consists of 7 sub-bars representing the proportions of the 7 macro-haplogroups H, HV*, I, J, K, T and U# in relevant cohort. There was no any difference in the proportion of seven macro-haplogroups in 15 cohorts ($P \approx 1$, Fisher exact test).

Supplementary Figure 3 Patients with PD with major sub-haplogroups of H have similar risk of progression to global cognitive impairment



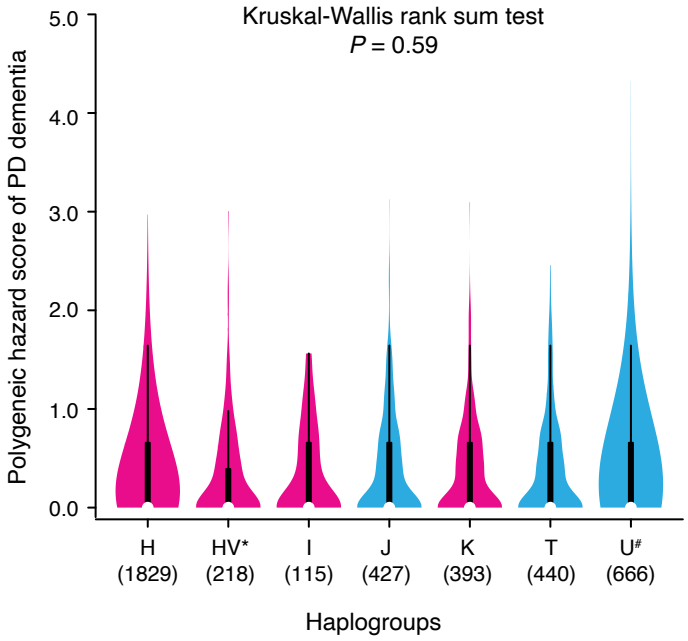
Cox regression analysis did not show any different hazard ratio (HR) to develop global cognitive impairment (MMSE $\leq$ 25) in combined population, according to the recommendation of the International Parkinson and Movement Disorder Society (MDS) Taks Force[29], among patients with PD in six major sub-haplogroups of H.

Supplementary Figure 4 Patients with PD in seven macro-haplogroups have similar risk of progression to HY3.



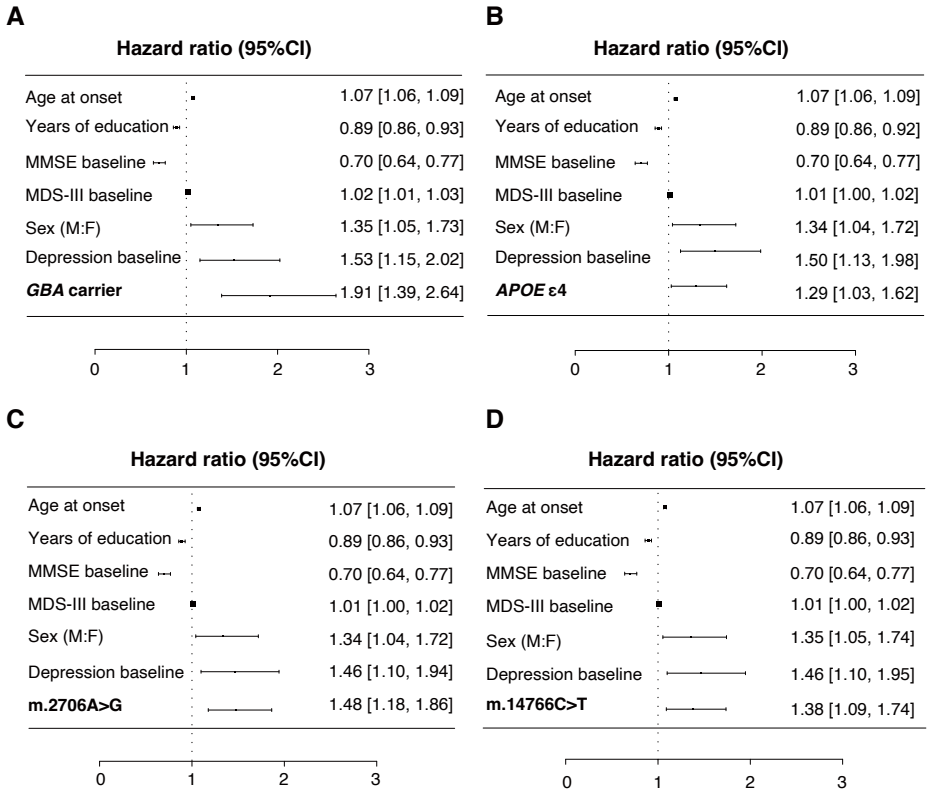
Cox regression analysis did not show any difference in hazard ratio (HR) for development of motor disability with postural instability (Hoehn & Yahr stage 3) during the progression of disease in seven macro-haplogroups from (**A**) discovery, (**B**) replication and (**C**) combined population.

Supplementary Figure 5 Patients with PD in seven mitochondrial macro-haplogroups have similar polygenic hazard scores



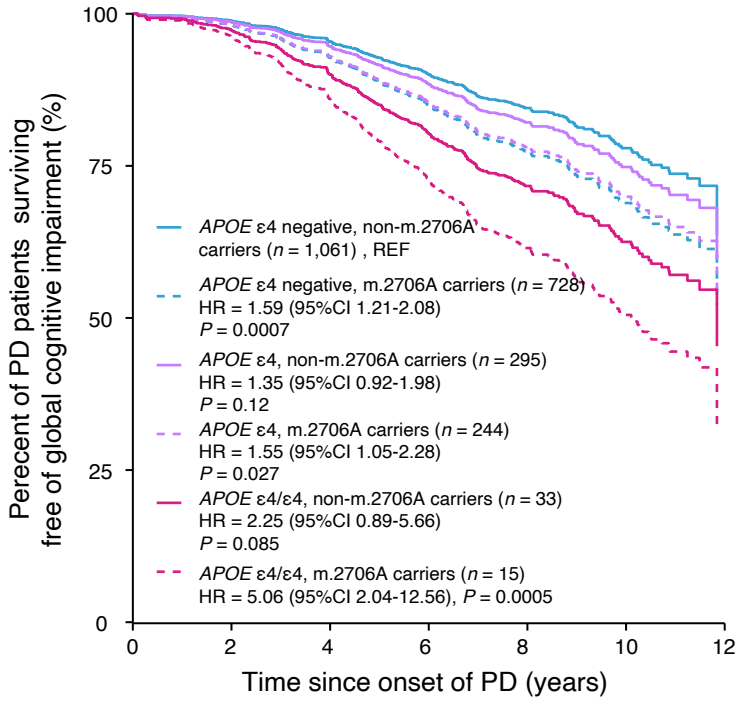
Violin-plot showed no significant difference between Polygenic hazard score to develop PD dementia among seven macro-haplogroups in combined population. Violin plot is a mixed of a box plot and a kernel density plot: the white dot represents the median, and black bar represents the interquartile range of score, the thin black line represents the rest of distribution and each side of the line is a kernel density estimation.

Supplementary Figure 6 The exploratory analysis for global cognitive impairment models with different genetic factors.



The forest plots show hazard ratios (Methods) for global cognitive impairment (GCI) in different genetic models (**A**) *GBA* carrier, (**B**) *APOE* ε4, (**C**) m.2706A>G and (**D**) m.14766C>T with the same six clinical risk factors. The squares represent point estimates, with the height of the square inversely proportional to the standard error of the estimates. The horizontal lines indicate 95% confidence intervals of the estimates.

Supplementary Figure 7 Exploratory analysis for global cognitive impairment models with *APOE* ε4 and m.2706A>G.



Covariate-adjusted survival curves for patients with PD stratified into six subgroups: *APOE* ε4 negative and non-m.2706A carriers (*n* = 1,061), *APOE* ε4 negative and m.2706A carriers (*n* = 728), *APOE* ε4 heterozygotes and non-m.2706A carriers (*n* = 295), *APOE* ε4 heterozygotes and m.2706A carriers (*n* = 244), *APOE* ε4/ε4 and non-m.2706A carriers (*n* = 33), *APOE* ε4/ε4 and m.2706A carriers (*n* = 15).

References

1. Locascio JJ, Eberly S, Liao Z*, et al*. Association between alpha-synuclein blood transcripts and early, neuroimaging-supported Parkinson's disease. *Brain*. Sep 2015;138(Pt 9):2659-71.

2. Writing Group for the NETiPDI, Kieburtz K, Tilley BC*, et al*. Effect of creatine monohydrate on clinical progression in patients with Parkinson disease: a randomized clinical trial. *Jama*. Feb 10 2015;313(6):584-93.

3. Corvol JC, Artaud F, Cormier-Dequaire F*, et al*. Longitudinal analysis of impulse control disorders in Parkinson disease. *Neurology*. Jul 17 2018;91(3):e189-e201.

4. Verbaan D, Marinus J, Visser M, van Rooden SM, Stiggelbout AM, van Hilten JJ. Patient-reported autonomic symptoms in Parkinson disease. *Neurology*. Jul 24 2007;69(4):333-41.

5. Williams-Gray CH, Evans JR, Goris A*, et al*. The distinct cognitive syndromes of Parkinson's disease: 5 year follow-up of the CamPaIGN cohort. *Brain*. Nov 2009;132(Pt 11):2958-69.

6. Williams-Gray CH, Mason SL, Evans JR*, et al*. The CamPaIGN study of Parkinson's disease: 10-year outlook in an incident population-based cohort. *Journal of neurology, neurosurgery, and psychiatry*. Nov 2013;84(11):1258-64.

7. Winder-Rhodes SE, Evans JR, Ban M*, et al*. Glucocerebrosidase mutations influence the natural history of Parkinson's disease in a community-based incident cohort. Research Support, Non-U.S. Gov't. *Brain*. Feb 2013;136(Pt 2):392-9.

8. Breen DP, Evans JR, Farrell K, Brayne C, Barker RA. Determinants of delayed diagnosis in Parkinson's disease. *Journal of neurology*. Aug 2013;260(8):1978-81.

9. Rosenthal LS, Drake D, Alcalay RN*, et al*. The NINDS Parkinson's disease biomarkers program. *Mov Disord*. Jun 2016;31(6):915-23.

10. Beach TG, Adler CH, Sue LI*, et al*. Arizona Study of Aging and Neurodegenerative Disorders and Brain and Body Donation Program. *Neuropathology : official journal of the Japanese Society of Neuropathology*. Aug 2015;35(4):354-89.

11. Alves G, Muller B, Herlofson K*, et al*. Incidence of Parkinson's disease in Norway: the Norwegian ParkWest study. *Journal of neurology, neurosurgery, and psychiatry*. Aug 2009;80(8):851-7.

12. Lucero C, Campbell MC, Flores H, Maiti B, Perlmutter JS, Foster ER. Cognitive reserve and beta-amyloid pathology in Parkinson disease. *Parkinsonism & related disorders*. Aug 2015;21(8):899-904.

13. Jankovic J, McDermott M, Carter J*, et al*. Variable expression of Parkinson's disease: a base-line analysis of the DATATOP cohort. The Parkinson Study Group. *Neurology*. Oct 1990;40(10):1529-34.

14. Ravina B, Tanner C, Dieuliis D*, et al*. A longitudinal program for biomarker development in Parkinson's disease: a feasibility study. *Mov Disord*. Oct 30 2009;24(14):2081-90.

15. Kadastik-Eerme L, Taba N, Asser T, Taba P. The increasing prevalence of Parkinson's disease in Estonia. *Acta Neurol Scand*. Sep 2018;138(3):251-258.

16. Parkinson Progression Marker I. The Parkinson Progression Marker Initiative (PPMI).

Progress in neurobiology. Dec 2011;95(4):629-35.

17. DATATOP: a multicenter controlled clinical trial in early Parkinson's disease. Parkinson Study Group. Clinical Trial
Comparative Study
Multicenter Study
Randomized Controlled Trial
Research Support, Non-U.S. Gov't
Research Support, U.S. Gov't, P.H.S. *Archives of neurology*. Oct 1989;46(10):1052-60.

18. Nalls MA, Pankratz N, Lill CM*, et al*. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nature genetics*. Sep 2014;46(9):989-93.

19. Hughes AJ, Daniel SE, Ben-Shlomo Y, Lees AJ. The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service. *Brain*. Apr 2002;125(Pt 4):861-70.

20. Goetz CG, Poewe W, Rascol O*, et al*. Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: status and recommendations. *Mov Disord*. Sep 2004;19(9):1020-8.

21. Hoops S, Nazem S, Siderowf AD*, et al*. Validity of the MoCA and MMSE in the detection of MCI and dementia in Parkinson disease. *Neurology*. Nov 24 2009;73(21):1738-45.

22. van Steenoven I, Aarsland D, Hurtig H*, et al*. Conversion between mini-mental state examination, montreal cognitive assessment, and dementia rating scale-2 scores in Parkinson's disease. *Mov Disord*. Dec 2014;29(14):1809-15.

23. Liu G, Peng J, Liao Z*, et al*. Genome-wide survival study identifies a novel synaptic locus and polygenic score for cognitive progression in Parkinson's disease. *Nature genetics*. May 6 2021;53(6):787-793.

24. Genomes Project C, Auton A, Brooks LD*, et al*. A global reference for human genetic variation. *Nature*. Oct 1 2015;526(7571):68-74.

25. Calabrese C, Simone D, Diroma MA*, et al*. MToolBox: a highly automated pipeline for heteroplasmy annotation and prioritization analysis of human mitochondrial variants in high-throughput sequencing. *Bioinformatics*. Nov 1 2014;30(21):3115-7.

26. Weissensteiner H, Pacher D, Kloss-Brandstatter A*, et al*. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res*. Jul 8 2016;44(W1):W58-63.

27. Helgason A, Hickey E, Goodacre S*, et al*. mtDNA and the islands of the North Atlantic: estimating the proportions of Norse and Gaelic ancestry. *Am J Hum Genet*. Mar 2001;68(3):723-37.

28. Stewart JB, Chinnery PF. The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nat Rev Genet*. Sep 2015;16(9):530-42.

29. Dubois B, Burn D, Goetz C*, et al*. Diagnostic procedures for Parkinson's disease dementia: recommendations from the movement disorder society task force. *Mov Disord*. Dec 2007;22(16):2314-24.

30. Liu G, Boot B, Locascio JJ*, et al*. Specifically neuropathic Gaucher's mutations accelerate cognitive decline in Parkinson's. *Annals of neurology*. Nov 2016;80(5):674-685.

31. Therneau TM. *A Package for Survival Analysis in S.* 2014.

32. Locascio JJ, Atri A. An overview of longitudinal data analysis methods for neurological research. *Dementia and geriatric cognitive disorders extra*. Jan 2011;1(1):330-57.

33. Liu G, Locascio JJ, Corvol JC*, et al*. Prediction of cognition in Parkinson's disease with a clinical-genetic score: a longitudinal analysis of nine cohorts. *The Lancet Neurology*. Aug 2017;16(8):620-629.

34. Pliss L, Tambets K, Loogvali EL*, et al*. Mitochondrial DNA portrait of Latvians: towards the understanding of the genetic structure of Baltic-speaking populations. *Ann Hum Genet*. Jul 2006;70(Pt 4):439-58.

35. Alvarez JC, Johnson DL, Lorente JA*, et al*. Characterization of human control region sequences for Spanish individuals in a forensic mtDNA data set. *Leg Med (Tokyo)*. Nov 2007;9(6):293-304.

36. Pereira L, Prata MJ, Amorim A. Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation. *Ann Hum Genet*. Nov 2000;64(Pt 6):491-506.

37. Dubut V, Chollet L, Murail P*, et al*. mtDNA polymorphisms in five French groups: importance of regional sampling. *Eur J Hum Genet*. Apr 2004;12(4):293-300.

38. Vidrová V, Tesařová M, Trefilová E, Honzík T, Magner M, Zeman J. Mitochondrial DNA Haplogroups in the Czech Population Compared to Other European Countries. *Human Biology*. 2008;80(6):669-674.

39. Messina F, Scorrano G, Labarga CM, Rolfo MF, Rickards O. Mitochondrial DNA variation in an isolated area of Central Italy. *Ann Hum Biol*. Jun 2010;37(3):385-402.

40. Maruszak A, Canter JA, Styczynska M, Zekanowski C, Barcikowska M. Mitochondrial haplogroup H and Alzheimer's disease--is there a connection? *Neurobiol Aging*. Nov 2009;30(11):1749-55.

41. Maruszak A, Safranow K, Branicki W*, et al*. The impact of mitochondrial and nuclear DNA variants on late-onset Alzheimer's disease risk. *J Alzheimers Dis*. 2011;27(1):197-210.

42. Carrieri G, Bonafe M, De Luca M*, et al*. Mitochondrial DNA haplogroups and APOE4 allele are non-independent variables in sporadic Alzheimer's disease. *Hum Genet*. Mar 2001;108(3):194-8.

43. Swerdlow RH, Hui D, Chalise P*, et al*. Exploratory analysis of mtDNA haplogroups in two Alzheimer's longitudinal cohorts. *Alzheimer's & Dementia*. 2020;16(8):1164-1172.

44. Ridge PG, Koop A, Maxwell TJ*, et al*. Mitochondrial haplotypes associated with biomarkers for Alzheimer's disease. *PLoS One*. 2013;8(9):e74158.

45. van der Walt JM, Dementieva YA, Martin ER*, et al*. Analysis of European mitochondrial haplogroups with Alzheimer disease risk. *Neuroscience Letters*. 2004;365(1):28-32.