medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Genome-wide association study of copy number variations in Parkinson's disease

Running title: Copy number variants impact on Parkinson's disease

Zied Landoulsi, PhD, [1,2]* Ashwin Ashok Kumar Sreelatha, MSC. M.Tech, [3] Claudia Schulte, MSC, [4,5] Dheeraj Reddy Bobbili, PhD, [1] Ludovica Montanucci, PhD, [6] Costin Leu, PhD, [7,8] Lisa-Marie Niestroj, PhD, [9] Emadeldin Hassanin, MSC, [1] Cloé Domenighetti, PhD, [10] Lukas Pavelka, MD, [2] Pierre-Emmanuel Sugier, PhD, [10] Milena Radivojkov-Blagojevic, MSc, [11] Peter Lichtner, PhD, [11] Berta Portugal, PhD, [12] Connor Edsall, PhD, [13] Jens Kru□ger, PhD, [14] Dena G Hernandez, PhD, [13] Cornelis Blauwendraat, PhD, [13] George D Mellick, PhD, [15] Alexander Zimprich, MD, [16] Walter Pirker, MD, [17] Manuela Tan, MSC, [18] Ekaterina Rogaeva, PhD, [19] Anthony E. Lang, MD, [19,20,21,22] Sulev Koks, MD, PhD, [23,24] Pille Taba, MD, PhD, [25,26] Suzanne Lesage, PhD, [27] Alexis Brice, MD, [27] Jean-Christophe Corvol, MD, PhD, [27,28] Marie-Christine Chartier-Harlin, PhD, [29] Eugenie Mutez, MD, PhD, [29] Kathrin Brockmann, MD [4,5] Angela B Deutschländer, MD, [30,31,32] Georges M Hadjigeorgiou, MD, [33,34] Efthimos Dardiotis, MD, [34] Leonidas Stefanis, MD, PhD, [35,36] Athina Maria Simitsi, MD, PhD, [36] Enza Maria Valente, MD, PHD [37,38] Simona Petrucci, MD, PhD, [39,40] Letizia Straniero, PhD, [41] Anna Zecchinelli, MD, [42] Gianni Pezzoli, MD, [42,43] Laura Brighina, MD, PhD, [44,45] Carlo Ferrarese, MD, PhD, [44,45] Grazia Annesi, PhD, [46] Andrea Quattrone, MD, PhD, [47] Monica Gagliardi, PhD, [48] Lena F Burbulla, PhD, [5,49,50,51] Hirotaka Matsuo, MD, PhD, [52] Akiyoshi Nakayama, MD, PhD, [52] Nobutaka Hattori, MD, PhD, [53] Kenya Nishioka, MD, PhD, [53] Sun Ju Chung, MD, PhD, [54] Yun Joong Kim, MD, PhD, [55] Pierre Kolber, MD, [56] Bart PC van de Warrenburg, MD, PhD, [57] Bastiaan R Bloem, MD, PhD, [57] Andrew B. Singleton, PhD, [13,58] Mathias Toft, MD, PhD, [59] Lasse Pihlstrom, MD, [59] Leonor Correia Guedes, MD, PhD, [60,61] Joaquim J Ferreira, MD, PhD, [60,62] Soraya Bardien, PhD, [63,64] Jonathan Carr, PhD, [65] Eduardo Tolosa, MD, PhD, [66,67] Mario Ezquerra, PhD, [68] Pau Pastor, MD, PhD, [69,70] Karin Wirdefeldt, MD, PhD, [71,72] Nancy L Pedersen, PhD, [72] Caroline Ran, PhD, [73] Andrea C Belin, PhD, [73] Andreas Puschmann, MD, PhD, [74] Carl E Clarke, MD, [75] Karen E Morrison, MD, [76] Dimitri Krainc, MD, PhD, [49] Matt J Farrer, PhD, [77] Dennis Lal, PhD, [6,78,79] Alexis Elbaz, PhD, [10] Thomas Gasser, MD, [4,5] Rejko Krüger, MD [1,2,56] Manu Sharma, PhD, [3*#] Patrick May, PhD, [1#] on behalf of the Comprehensive Unbiased Risk Factor Assessment for Genetics and Environment in Parkinson's Disease (COURAGE-PD) consortium.

[#]Equal author contribution.

1 Luxembourg Centre for Systems Biomedicine, University of Luxembourg; L-4367, Esch-sur-Alzette, Luxembourg

2 Transversal Translational Medicine, Luxembourg Institute of Health, Strassen, Luxembourg

3 Centre for Genetic Epidemiology, Institute for Clinical Epidemiology and Applied Biometry, University of Tubingen, Germany

4 Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tubingen, Germany.

5 German Center for Neurodegenerative Diseases (DZNE), Tubingen, Germany.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

6 Department of Neurology, McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, TX, USA

7 Department of Clinical and Experimental Epilepsy, Institute of Neurology, University College London, London, UK

8 Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH, USA

9 Cologne Center for Genomics (CCG), Medical Faculty of the University of Cologne, Cologne, Germany

10 Université Paris-Saclay, UVSQ, Inserm, Gustave Roussy, CESP, 94805 , Villejuif, France.

11 Institute of Human Genetics, Helmholtz Zentrum Munchen, Neuherberg, Germany.

12 Department of Precision Health, Luxembourg Institute of Health, Strassen, Luxembourg.

13 Molecular Genetics Section, Laboratory of Neurogenetics, NIA, NIH, Bethesda, MD 20892, USA.

14 Group of Applied Bioinformatics, University of Tubingen, Tubingen, Germany.

15 Griffith Institute for Drug Discovery, Griffith University, Don Young Road, Nathan, Queensland, Australia.

16 Department of Neurology, Medical University of Vienna, Austria.

17 Department of Neurology, Klinik Ottakring, Vienna Austria

18 Department of Clinical and Movement Neurosciences, UCL Queen Square Institute of Neurology, University College London, London, UK.

19 Tanz Centre for Research in Neurodegenerative Diseases, University of Toronto, Toronto, Ontario, Canada.

20 Edmond J. Safra Program in Parkinson's Disease, Morton and Gloria Shulman Movement Disorders Clinic, Toronto Western Hospital, UHN, Toronto, Ontario, Canada.

21 Division of Neurology, University of Toronto, Toronto, Ontario, Canada

22 Krembil Brain Institute, Toronto, Ontario, Canada.

23 Centre for Molecular Medicine and Innovative Therapeutics, Murdoch University, Murdoch, Australia.

24 Perron Institute for Neurological and Translational Science, Nedlands, Western Australia, Australia.

25 Department of Neurology and Neurosurgery, University of Tartu, Estonia.

26 Neurology Clinic, Tartu University Hospital, Tartu, Estonia.

27 Sorbonne Université, Paris Brain Institute – ICM, Inserm, CNRS, Paris, France

28 Assistance Publique Hôpitaux de Paris, Department of Neurology, CIC Neurosciences, Pitié-Salpêtrière Hospital, Paris, France.

29 Univ. Lille, Inserm, CHU Lille, UMR-S 1172 - JPArc - Centre de Recherche Lille Neurosciences & Cognition, F-59000 Lille, France.

30 Department of Neurology, Ludwig Maximilians University of Munich, Germany.

31 Department of Neurology, Max Planck Institute of Psychiatry, Munich, Germany.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

32 Department of Neurology and Department of Clinical Genomics, Mayo Clinic Florida, Jacksonville, FL, USA.

33 Department of Neurology, Medical School, University of Cyprus, Nicosia, Cyprus.

34 Department of Neurology, Laboratory of Neurogenetics, University of Thessaly, University Hospital of Larissa, Larissa, Greece.

35 Center of Clinical Research, Experimental Surgery and Translational Research, Biomedical Research Foundation of the Academy of Athens, Athens, Greece.

36 1st Department of Neurology, Eginition Hospital, Medical School, National and Kapodistrian University of Athens, Athens, Greece.

37 Department of Molecular Medicine, University of Pavia, Pavia, Italy.

38 Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Mondino Foundation, Pavia, Italy.

39 UOC Medical Genetics and Advanced Cell Diagnostics, S. Andrea University Hospital, Rome, Italy.

40 Department of Clinical and Molecular Medicine, Sapienza University of Rome, Italy

UOC Medical Genetics and Advanced Cell Diagnostics, S. Andrea University Hospital, Rome, Italy.

41 Department of Biomedical Sciences - Humanitas University, Milan, Italy.

42 Parkinson Institute, Azienda Socio Sanitaria Territoriale (ASST) Gaetano Pini/CTO, Milano, Italia.

43 Fondazione Grigioni per il Morbo di Parkinson, Milan, Italy.

44 Department of Neurology, San Gerardo Hospital, Milan, Italy.

45 Center for Neuroscience, University of Milano Bicocca, Monza, Italy.

46 Institute for Biomedical Research and Innovation, National Research Council, Cosenza, Italy

47 Institute of Neurology, Department of Medical and Surgical Sciences, Magna Graecia University, Catanzaro, Italy.

48 Department of Medical and Surgical Sciences, Neuroscience Research Center, Magna Graecia University, Catanzaro, Italy

49 Department of Neurology, Northwestern University Feinberg School of Medicine, Chicago, Illinois 60611, United States.

50 Metabolic Biochemistry, Biomedical Center (BMC), Faculty of Medicine, Ludwig-Maximilians University, Munich, Germany.

51 Munich Cluster for Systems Neurology (SyNergy), Munich, Germany.

52 Department of Integrative Physiology and Bio-Nano Medicine, National Defense Medical College, Saitama 359-8513, Japan.

53 Department of Neurology, Juntendo University School of Medicine, Bunkyo-ku, Tokyo 113-8421, Japan.

54 Department of Neurology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea.

55 Department of Neurology, Yonsei University College of Medicine, Seoul, South Korea.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

56 Centre Hospitalier du Luxembourg, Parkinson Research Clinic, Luxembourg, Luxembourg

57 Radboud University Medical Centre, Donders Institute for Brain, Cognition and Behaviour, Department of Neurology, Nijmegen, The Netherlands.

58 Center For Alzheimer's and Related Dementias, NIA, NIH, Bethesda, MD, USA.

59 Department of Neurology, Oslo University Hospital, Oslo, Norway.

60 Instituto de Medicina Molecular João Lobo Antunes, Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal.

61 Department of Neurosciences and Mental Health, Neurology, Hospital de Santa Maria, Centro Hospitalar Universitario Lisboa Norte (CHULN), Lisbon, Portugal.

62 Laboratory of Clinical Pharmacology and Therapeutics, Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal.

63 Division of Molecular Biology and Human Genetics, Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Stellenbosch University, South Africa.

64 South African Medical Research Council / Stellenbosch University Genomics of Brain Disorders Research Unit, Stellenbosch University, Cape Town, South Africa

65 Division of Neurology, Department of Medicine, Faculty of Medicine and Health Sciences, Stellenbosch University, South Africa.

66 Parkinson's disease &Movement Disorders Unit, Neurology Service, Hospital Clínic de Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), University of Barcelona, Barcelona, Spain.

67 Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED: CB06/05/0018-ISCIII) Barcelona, Spain.

68 Lab of Parkinson Disease and Other Neurodegenerative Movement Disorders, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Institut de Neurociències, Universitat de Barcelona, ES-08036 Barcelona, Catalonia.

69 Fundació per la Recerca Biomèdica i Social Mútua Terrassa, Terrassa, Barcelona, Spain.

70 Movement Disorders Unit, Department of Neurology, Hospital Universitari Mutua de Terrassa, Terrassa, Barcelona, Spain.

71 Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden.

72 Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.

73 Department of Neuroscience, Karolinska Institutet, Stockholm, Sweden.

74 Lund University, Skåne University Hospital, Department of Clinical Sciences Lund, Neurology, Getingevägen 4, 221 85, Lund, Sweden.

75 University of Birmingham and Sandwell and West Birmingham Hospitals NHS Trust, United Kingdom.

76 Faculty of Medicine, Health and Life Sciences, Queens University, Belfast, United Kingdom.

77 Department of Neurology, McKnight Brain Institute, University of Florida, Gainesville, FL, USA.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

78 Epilepsy Center, Neurological Institute, Cleveland Clinic, Cleveland, Ohio, USA.

79 Stanley Center for Psychiatric Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.

* Corresponding authors: Zied Landoulsi zied.landoulsi@lih.lu; Manu Sharma manu.sharma@uni-tuebingen.de

medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

Objective:

Our study investigates the impact of copy number variations (CNVs) on Parkinson's disease (PD) pathogenesis using genome-wide data, aiming to uncover novel genetic mechanisms and improve the understanding of the role of CNVs in sporadic PD.

Methods:

We applied a sliding window approach to perform CNV-GWAS and conducted genome-wide burden analyses on CNV data from 11,035 PD patients (including 2,731 early-onset PD (EOPD)) and 8,901 controls from the COURAGE-PD consortium.

Results:

We identified 14 genome-wide significant CNV loci associated with PD, including one deletion and 13 duplications. Among these, duplications in 7q22.1, 11q12.3 and 7q33 displayed the highest effect. Two significant duplications overlapped with PD-related genes *SNCA* and *VPS13C*, but none overlapped with recent significant SNP-based GWAS findings. Five duplications included genes associated with neurological disease, and four overlapping genes were dosage-sensitive and intolerant to loss-of-function variants. Enriched pathways included neurodegeneration, steroid hormone biosynthesis, and lipid metabolism. In early-onset cases, four loci were significantly associated with EOPD, including three known duplications and one novel deletion in *PRKN*. CNV burden analysis showed a higher prevalence of CNVs in PD-related genes in patients compared to controls (OR=1.56 [1.18-2.09], p=0.0013), with *PRKN* showing the highest burden (OR=1.47 [1.10-1.98], p=0.026). Patients with CNVs in *PRKN* had an earlier disease onset. Burden analysis with controls and EOPD patients showed similar results.

Interpretation:

This is the largest CNV-based GWAS in PD identifying novel CNV regions and confirming the significant CNV burden in EOPD, primarily driven by the *PRKN* gene, warranting further investigation.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Introduction

A global survey estimates that Parkinson's disease (PD) is the world's fastest-growing neurodegenerative disorder, surpassing even Alzheimer's disease [1]. Concerted efforts are required to unravel the underlying complexity of a disease in which genetics and environmental factors appear to play an important role [2,3]. To date, considerable progress has been made in understanding the genetic basis of PD by identifying families in which the disease segregates in Mendelian fashion and by applying array-based approaches (commonly used in genome-wide association studies (GWAS)) to identify risk factors for sporadic forms of PD [4,5].

GWAS have been successful in identifying a number of loci that are potentially relevant for PD [6,7]. While this success has been remarkable, the genetic variability explained so far is between 19-39% [6] indicating that the genetic variability in the form of chromosomal arrangements - duplications, deletions - commonly referred to as copy number variations (CNVs) - may be an important player in explaining the genetic variability in PD [8–11]. Indeed, previously studies identified several CNVs in PD-related genes. For example, genomic multiplications in the *SNCA* gene were shown to cause familial [12] and sporadic [13] forms of PD. Duplication or triplication of this gene can result in an excess of alpha-synuclein, which may then aggregate and form Lewy bodies, ultimately contributing to neurodegeneration observed in PD [10]. Furthermore, CNVs in *PRKN* are the most prevalent CNVs among known PD genes [14–16], owing to its location in one of the most mutation-susceptible regions of the human genome [17]. Deletions in familial PD forms were previously described in *PINK1* [18,19] and *PARK7* [20,21], but are less common than in *PRKN*. Unlike *SNCA*, deletions in homozygous-driven early-onset PD forms have demonstrated the role of loss of function genes in the etiology of the disease, while the pathogenic role of single heterozygous CNVs is still controversial.

The above examples support the role of CNVs in genes that are *bona-fide* loci for monogenetic PD. In contrast, the role and impact of CNVs on sporadic PD is still unclear. However, previous studies have suggested a potential role of CNVs in sporadic PD as well. For instance, a genome-wide CNV burden analysis in a Latin American cohort [22] showed that PD patients were significantly enriched only for CNVs affecting known PD genes, but they could not identify any putative CNV in PD pathogenesis at the genome-wide level. Their failure to identify putative CNVs could be attributed to the small sample size. Similarly, another study could only validate that the CNV within the *PRKN* locus was significantly associated with PD susceptibility [23]. Thus, a CNV-based GWAS in large, well-powered and well characterized PD cohorts may reveal novel molecular pathways associated with the disease, potentially advancing efforts to understand the role of CNVs in PD pathogenesis.

The Comprehensive Unbiased Risk Factor Assessment for Genetics and Environment in Parkinson's Disease (COURAGE-PD) is a worldwide collaboration consortium, aimed at understanding the roles of genetics and environment in PD [7,24,25]. In the present study, we leveraged the genome-wide data to understand the impact of CNVs on PD in COURAGE-PD. Given the fact that there is no consensus on the best method for detecting or analyzing genome-wide significant CNVs [9], we applied the sliding

medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

window approach to perform CNV-GWAS, and employed genome-wide burden analyses to identify novel genome-wide significant CNV regions and to investigate their impact on the disease.

Subjects and Methods

Study cohort

The COURAGE-PD consortium includes data from 15,849 PD patients and 11,444 controls of predominantly European ancestry from 35 cohorts [7]. In our study, we used genotyping data from 22,329 individuals from 25 European ancestry cohorts originating from 15 European countries. Genotyping quality control (QC) was conducted independently for each cohort, by following the standard procedure previously reported [7]. Patients with early-onset PD (EOPD) were defined as those diagnosed before the age of 50 (age at onset (AAO) <= 50 years) [26].

Copy number variant calling and quality control

We created a custom population B-allele frequency (BAF) and GC wave-adjusted log R ratio (LRR) intensity file using GenomeStudio (v2.0.5 Illumina) for all the samples that passed genotyping QC and employed PennCNV (v1.0.5, [27]) to detect CNVs in our dataset. The analysis was restricted to autosomal CNVs, as calls from the sex chromosomes are often of poor quality [27]. We used a post-CNV calling QC pipeline including standard parameters as previously described in studies on CNV calling from SNP array data for PD [22] or epilepsy [28,29]. As a first step, adjacent CNV calls were merged into a single call if the number of overlapping markers between them was less than 20% of the total number when both segments were combined. This was followed by intensity-based QC to exclude samples with low-quality data. Samples with a log R Ratio (LRR) standard deviation of less than 0.24, an absolute value of the waviness factor of less than 0.03, and a B-allele frequency (BAF) drift of less than 0.001 were retained. These thresholds correspond to the median plus 3 SDs. Furthermore, CNV calls with more than 50% overlap with known problematic genomic regions [27], including centromeric, telomeric, and HLA regions, were also excluded before analysis. Next, we removed CNVs that met the following criteria: they spanned fewer than 20 SNPs, were less than 20 kilobases (kb) in length, and had an SNP density of less than 0.0001 (number of SNPs/length of CNV). For CNVs spanning at least 20 SNPs and longer than 1 Mb, SNP density was not considered. Finally, we applied a series of filters to identify rare CNVs. The first step was to assign a specific frequency count to each CNV call using PLINK v.1.07 [30]. This was followed by applying a filter to exclude common CNVs, retaining only rare variants that overlapped with CNVs in at least 1% of all samples. In the second step, CNVs with an overlap of at least 50% with reported common CNVs (allele frequency >1%) in the Database of Genomic Variants (DGV Gold standard dataset [31]) and DECIPHER population CNVs frequencies [32] were excluded. The filtered rare CNVs were annotated

medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

for gene content using refGene, including the gene name and corresponding coordinates in the hg19 assembly with ANNOVAR (v 2020-06-08).

Sliding windows CNV analysis and assessment of genome-wide significance

A segment-based rare CNV burden analysis was performed to identify genomic regions with a significant increase in rare CNVs in PD cases compared to controls. This analysis was conducted separately for each type of CNV (deletion or duplication) using a sliding window approach [33]. The sliding windows model allowed association testing of all autosomes through 267,237 sliding windows characterized by a window size of 200 kb and a step size of 10 kb, corresponding to 13,339.6 non-overlapping windows. The threshold for genome-wide significance was set to $\alpha = 3.74 \times 10e\text{-}06$ after Bonferroni correction for multiple testing. For each of the genomic regions, the number of overlapping CNVs was counted separately for cases and controls for deletions or duplications. We considered a minimum overlap of 10% between the CNV and the genomic window to identify the potential burden of small deletions or duplications ($\geq$ 20 kb). The one-sided Fisher's exact test was used as a test statistic for the CNVs collapsed for each segment. The analysis was performed using the rCNV docker (https://hub.docker.com/r/talkowski/rcnv) [33], and custom Python (version 3.7.9) and R (version 4.3.1) scripts. To assess the impact of age at PD onset, the same analysis was subsequently stratified based on EOPD, encompassing all the control subjects and patients with EOPD.

Association fine-mapping

The observation that a considerable number of large rare CNVs involve the deletion or duplication of multiple adjacent genes, led us to hypothesize that the most associated genes were not causal, but rather gained significance due to proximity to true causal genes. This is analogous to the linkage disequilibrium effect observed in GWAS of common variants. To address this issue, Collins *et al.* employed a Bayesian fine-mapping algorithm to define the 95% credible set of causal elements or genes at each genome-wide significant locus for deletions and duplications [33]. This method prioritizes the most probable causal genes based on their association statistics. The Bayesian algorithm was employed to calculate the approximate Bayes factor (ABF) for each window, as previously described by Wakefield [34]. The ABF offers an alternative to the P-values for assessing the significance of association by providing a summary measure that ranks these associations. Bayesian model averaging was used to estimate the null variance of true causal loci across taking into account prior information regarding the mean of all significant windows and the most significant window per block. The minimal set of windows that constitutes the 95% credible set for each block was defined by ranking the windows in descending order according to their ABF.

Predictive scores and gene annotation

medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

In order to assess the functional impact of CNVs, the spanned genes were annotated using three genetic constraint scores that are available in gnomAD (https://gnomad.broadinstitute.org): the Loss-of-function Observed/Expected Upper Bound Fraction (LOEUF), the Predicted Probability of Haploinsufficiency (pHaplo) and the Predicted Probability of Triplosensitivity (pTriplo). LOEUF is a quantitative measure of the observed depletion (or enrichment) of loss-of-function variants in gnomAD compared to a null mutation model. Values are ranked from 0 with no theoretical maximum. Genes with smaller values (closer to zero) are more intolerant to mutations [35].The two scores pHaplo and pTriplo reflect the likelihood that whole-copy deletion or duplication, respectively, of each gene would be enriched in a cohort of individuals affected by severe, early-onset diseases as compared to the general population [33]. We considered genes in the most constrained decile, defined as those with a LOEUF score of 0.268 or below. A pHaplo score ≥ 0.86 and a pTriplo score ≥ 0.94 indicate that the average effect sizes of deletions/duplications are comparable to the loss-of-function of genes typically restricted against protein-truncating variants, with an odds ratio (OR) ≥ 2.7. Similarly, a pHaplo score ≥ 0.55 and a pTriplo score≥ 0.68 indicate an OR ≥ 2.

Enriched biological pathways, diseases, and tissues expression

All genes within credible intervals underwent pathway enrichment using the EnrichR package [36] using Gene Ontology (GO) terms, Kyoto Encyclopedia of Genes and Genomes (KEGG), Reactome and DisGeNET pathways [37] databases. The Functional Mapping and Annotation of GWAS (FUMA version 1.5.2; https://fuma.ctglab.nl) web server was used to investigate tissue expression enrichment based on the GTEx v8 54 tissue types database. The gene-set enrichment results were subjected to correction for multiple testing using the Bonferroni method.

Genome-wide burden and survival analysis

The burden of rare CNVs associated with PD was calculated using distinct categories to determine their relative impact on PD risk, as previously reported [22]. These categories included: (1) carrier status of genome-wide CNV burden, including CNVs in non-genic regions, for all CNVs, not distinguishing between deletions and duplications (2) carrier status of any exonic or intronic CNVs intersecting with 'any gene' except those associated with PD, (3) carrier status of CNVs intersecting with exonic or intronic regions of the six major ‘PD-related genes’ according to MDS gene classification (https://www.mdsgene.org): *LRRK2*, *SNCA*, *VPS35* for dominant forms of classical parkinsonism and *PRKN*, *PARK7*, and *PINK1* for recessive forms of early-onset parkinsonism, (4) carrier status of CNVs intersecting only with exonic regions of the ‘PD-related genes’, (5) carrier status of CNVs in *PRKN* only and (6) carrier status of large CNVs (≥1 Mb in length). To compare the CNV burden between PD patients and controls, we used the *glm* function in R (v4.3.1) to fit a logistic regression model. This allowed the calculation of the OR with a 95% confidence interval and p-values. We used

medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

sex, age at assessment, and the first five principal components (PCs) from the population stratification as covariates for the regression model. Cox proportional hazards regression analyses and Kaplan–Meier curves were generated using the *survival* R package [38] with age defined as age at last visit for controls and AAO for cases. Controls were included as censored observations given that it was only known that they did not develop PD up to their last visit. Hazard ratios (HRs) and 95% confidence intervals (CIs) of PD were estimated by Cox proportional hazards models. Sex and the first five PCs were included as covariates. All the P-values underwent adjustment using the Bonferroni method to correct for multiple testing.

Results

COURAGE-PD cohort

The final dataset for the COURAGE-PD cohort comprised 11,035 patients with PD and 8,901 control, all of European descent, following the QC steps for the genotyping data. The demographic characteristics of each cohort are presented in Supplementary Table 1. A total of 2,731 PD patients (24.7% of the total patients) were identified as having EOPD (mean AAO 43.1±6.1 and mean age at assessment 54.4±9.3 years). We initially detected 1,098,221 CNVs in a total of 10,877 PD patients and 8,534 controls. After all QC and filtering steps (see Methods), the final number of rare CNVs was 28,263, including 7,843 duplications and 20,420 deletions in 3,896 PD patients and 3,299 controls. CNV analysis showed that 36.0% of the samples carried at least one QC-passed CNV. The characteristics of our CNV analysis are shown in Supplementary Table 2.

Identification of 14 genome-wide significant PD-associated CNV regions

A total of 267,237 genomic segments of 200 kb size in a 10 kb sliding window approach were scanned [33]. After adjusting for the multiple testing and applying the fine-mapping approach, we identified 14 genome-wide significant CNV loci associated with PD. The 14 loci comprised one deletion (210 kb) and 13 duplications (size range: 220 to 560 kb, Fig 1 and Table 1). The genome-wide deletion identified in this study was a single-copy loss and spanned two genes *COL18A1, and POFUT2* (Table 1, Supplementary Table 5). All duplications, except for the two-copy gain in 8q23.3, were single-copy gains. They spanned a total of 68 genes, with three duplications in 7q22.1, 11q12.3, and 7q33 showing the highest ORs (Table 1, Supplementary Table 5). Of note, a nearly significant deletion locus in 5q32 was observed, with an adjusted p-value of 3.76e-06 (slightly higher than the adjusted p-value threshold of ≤ 3.7410e-06). This locus encompasses the *STK32A*, *DPYSL3*, and *JAKMIP2* genes.

Of the 14 genome-wide significant CNV regions identified in this study, none overlapped with the most recent SNP-based GWAS study [6]. However, two genome-wide significant CNV regions overlapped with two PD-related genes, namely *SNCA* and *VPS13C* (Table 1). The 4q22.1 duplication

medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

interval encompasses exons 1 to 4 of *SNCA* and the 15q22.2 duplication interval overlapped with the entire coding region of *VPS13C*. These two rare duplication events did not overlap with the common significant GWAS SNPs identified in *SNCA* and *VPS13C* [6]. Furthermore, the correlation between these SNPs and the closest duplication SNP demonstrated that these loci are in complete linkage equilibrium ($r^2 = 0$).

Interestingly, five of the identified duplication regions encompass genes associated with neurological diseases (Table 1). For instance, the 1q42.12 region, which includes the *DNAH14* gene, is associated with neurodevelopmental disorder. The 6q21 duplication region encompasses the *FIG4* gene, which has been linked to Charcot-Marie-Tooth (CMT) and amyotrophic lateral sclerosis (ALS). Similarly, the *ATXN2* gene at 12q24.12 is associated with spinocerebellar ataxia and may also be linked to ALS and late-onset PD. Furthermore, *CLN5* at 13q22.3 is responsible for neuronal ceroid lipofuscinosis, while *FBXL3* is linked to a form of intellectual developmental disorder. Additionally, the 19p13.11 duplication region contains two genes associated with immunodeficiency: *FCHO1* and *JAK3*.

The next step was to assess the susceptibility of these genes to be dosage sensitive and their intolerance to loss-of-function variants. The LOEUF score for five genes located in duplication intervals was less than 0.268, including *ATXN2* (12q24.12), *MYCBP2* (13q22.3), *FCHO1* (19p13.11), *UNC13A* (19p13.11) and *ATRN* (20p13). This indicates a greater likelihood of dosage sensitivity (see Methods section and Supplementary Table 5). Among duplications, 14 genes are likely to be located in a region of the genome that is intolerant to duplication, with a pTriplo score greater than 0.68 (of which seven genes have a pTriplo score greater than 0.94, as detailed in the Methods section and Supplementary Table 5). The 20p13 interval was found to span the majority of these genes (n = 6), with two genes located in the 13q22.3 interval. In total, four genes were predicted to be dosage sensitive, either with LOEUF or pTriplo score: *ATXN2* (12q24.12), *MYCBP2* (13q22.3), *UNC13A* (19p13.11) and *ATRN* (20p13). In the nearly-significant 5q32 deletion locus, the *DPYSL3* and *JAKMIP2* genes are annotated as highly dose sensitive (LOEUF scores of 0.22 and 0.18, pHaplo of 0.97 and 0.92 respectively).

Identification of four genome-wide significant EOPD-associated CNV regions

To gain further insight into the genome-wide landscape of EOPD, we conducted a subset analysis comparing 2,731 PD patients with an AAO less than 50 years old and controls. We identified four genome-wide loci that were significantly associated with EOPD (Fig 2 and Table 2). Of note, three of these (1q42.12, 6q21, and 11q12.3, Fig 1 and Table 1) were already identified as duplications in all the patients. Furthermore, a new deletion region (590 kb in 6q21) was identified, covering the exon 2-6 of the PD-related *PRKN* gene.

Enrichment analysis

medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

We applied enrichment analysis to identify significantly enriched pathways in the GO, KEGG, and Reactome databases considering all CNVs overlapping genes within the credible intervals (70 genes overlapping duplications and two genes overlapping deletions, as detailed in Table 1). The top 10 most significant enriched pathways for each database are shown in Figure 3. Our analysis identified several significantly enriched pathways, which can be classified into the following categories: (i) steroid hormone biosynthesis (Estrogen 16-Alpha-Hydroxylase Activity, (GO:0101020), Steroid Hydroxylase Activity (GO:0008395)), (ii) several metabolic processes including lipid and phospholipid, (iii) enzymatic activities (Phospholipase and N-acetyltransferase), (iv) drug metabolism and (v) others. Among the other terms, two KEGG pathways are worthy of particular mention: 'Pathways of Neurodegeneration' and 'Lysosome' (Fig 3B). Neurological disorders, particularly movement disorders (e.g., restless legs syndrome and fatigable weakness of respiratory/swallowing muscles), were among the significant enriched diseases and traits from DisGeNET (Fig 3D). The DisGeNet term "Parkinsonian tremor" was also found to be significant, with the two PD-related genes *SNCA* and *VPS13C* being the only ones to emerge. We noted that the removal of the two deletion genes (*COL18A1*, *POFUT2*) did not affect the enrichment result displayed in Figure 3, indicating that the enrichment result is influenced mainly by duplicated genes. Next, we used FUMA to test whether the candidate genes were enriched for expression in 54 tissues. We did not find any significant expression levels in tissues after multiple-testing correction (Supplementary Fig 1). The top-ranked three tissues were the liver, adipose subcutaneous tissue, and the adrenal gland.

CNV burden analysis

We used logistic regression to compare the rare CNV burden on a genome-wide level in PD patients and controls across the predefined categories (see Methods). No significant difference was observed for genome-wide CNV burden (OR=1.04 [0.98-1.11], adjusted p-value (p_{adj})=0.3), duplications (OR=1.09 [1.01-1.17], p_{adj} = 0.46), or deletions (OR=1.00 [0.93-1.08], p_{adj}=0.89, Fig 4A), implicating that cumulative burden of CNVs does not alter the risk of getting PD. The same results were observed for the CNVs burden for all the genes excluding PD-related ones (OR=1.02 [0.96-1.09], p_{adj}=0.56) and large CNVs exceeding 1Mb in length (OR=0.97 [0.81-1.17], p_{adj}=0.89, Fig 4A).

In contrast, we observed a significant CNV burden for PD-related genes when considering all CNVs (OR=1.56 [1.18-2.09], p_{adj}=0.0013) with a slightly more pronounced association for CNVs overlapping exonic regions (OR=1.64 [1.18-2.30], p_{adj}=0.013, Fig 4A). CNVs in *PRKN* were the primary contributors to this finding (OR=1.47 [1.10-1.98], p_{adj}=0.026, Fig 4A). Indeed, among the 229 individuals carrying CNVs in PD-related genes, 91.7% were carriers of *PRKN* CNVs (summarised in Supplementary Table 3 and detailed in Supplementary Table 4). All these *PRKN* CNVs mostly consisted of one-copy duplications or duplications, except for seven homozygous deletions in PD patients (Table S3-4). CNVs in *PRKN* were identified in 135 PD patients (115 exonic and 36 intronic, representing 1.2% of the total patients) and 75 controls (56 exonic and 22 intronic,

medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

representing 0.8% of controls, Table S3-4). The AAO for patients with CNVs in *PRKN* was significantly earlier than that of patients without CNVs in *PRKN* (52.8±11.5 vs. 58.7±15.0 years, Mann-Whitney test p-value = 2.5e-05). Moreover, patients carrying exonic *PRKN* variants exhibited a significantly earlier AAO (50.8±15.5 years) compared to those carrying intronic variants (58.0 ± 12.2 years, Mann-Whitney test p-value = 0.02).

A subset of samples from one of the COURAGE-PD sites (Gasser/Sharma study, see Supplementary table 1) were previously screened for CNVs in *PRKN* and *SNCA* using the multiplex ligation-dependent probe amplification (MLPA) assay [39]. Out of 26 individuals with *PRKN* CNVs (21 patients and 5 controls), MLPA confirmed the CNVs in 22 individuals, and one patient with *SNCA* CNV was confirmed by MLPA (Supplementary Table S4), yielding a confirmation rate of 85%.

EOPD CNV burden analysis

We also performed a subset analysis comparing only the 2,731 cases with AAO ≤ 50 years with controls to evaluate the rare CNV burden in patients with EOPD. Again, our findings revealed a significant increase in CNVs in PD-related genes (*OR*=2.43 [1.64-3.59], p_{adj}=2.8e-05) with a slightly stronger association when examining CNVs that intersected with exonic regions (*OR*=3.10 [2.00-4.81], p_{adj}=3.2e-06, Fig 4A). Among the EOPD cohort, 62 patients (2.2% of EOPD patients) were identified as carriers of CNVs in PD-related genes, with *PRKN* being the most common (*OR*=2.32 [1.54-3.46], $p\text{-}value_{adj}$=1.1e-04, Fig 4 A), found in 57 patients (2.1% of EOPD patients), encompassing 48 exonic and 7 intronic CNVs. Carriers of the seven homozygous *PRKN* deletions mentioned above were EOPD patients.

Survival analysis

Kaplan–Meier analysis revealed that individuals with a CNV in PD-related genes had significantly earlier symptom onset compared to those carrying CNVs in other genes or non-carriers (log-rank test, p-value = 7.0 e-06; Fig. 4. B). Additionally, a Cox proportional hazards regression, adjusted for sex, and the first five ancestry PCs, indicated that having a CNV in PD-related genes was associated with a significantly increased hazard of earlier AAO (HR=1.48 [0.1-4.8]; *p-value* = 1.2e-06, Fig 4B). We further evaluated AAO exclusively in PD patients, comparing those with a CNV in PD-related genes to patients with other or no CNVs. The analysis also indicated that carrying at least one CNV leads to an earlier onset of symptoms (HR=1.39 [0.1-4.0]; p-value = 4.4e-05, Fig 4C).

Discussion

CNVs have long been established as a major contributor to disease risk in a wide range of complex diseases including epilepsy, neuropsychiatric and neurodegenerative disorders [10,22,28,29]. However, no large-scale study has been conducted at the genome-wide level to decipher the role of CNVs in

medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

sporadic PD. Thus, the current study aimed to extend the knowledge of potential role and impact of CNVs on PD. To achieve this, we employed a well-characterized multinational PD controls cohort from the COURAGE-PD consortium, leveraging a cohort of 19,936 individuals of European descent. Employing a rigorous CNV calling, quality control, and filtering pipeline, we systematically screened the genome to identify potential novel significant CNV regions for PD using the sliding window approach [33]. This was followed by an unbiased assessment to evaluate the CNV burden across various categories. Given the lack of consensus on optimal methods for detecting and analysing genome-wide significant CNVs [9], we employed the recently developed sliding window method already used to generate a genome-wide dosage sensitivity catalogue across 54 disorders, identifying 163 dosage-sensitive segments [33]. Application of such a method to research in epilepsy identified seizure-associated loci that had previously been demonstrated to be genome-wide significant in a previous study [29]. The GWAS-based CNV analysis led for the first time to the identification of novel CNV regions including one deletion and 13 duplication regions associated with PD risk. Importantly, our analysis identified two CNV regions that overlapped with previously reported PD-related genes, namely *SNCA* and *VPS13C*, demonstrating the robustness of our sliding window approach. Duplication or triplication of *SNCA* can result in an excess of alpha-synuclein, which may then aggregate and form Lewy bodies, ultimately contributing to neurodegeneration observed in PD [12,40,41]. A few rare homozygous variants in *VPS13C* have been identified as a cause of familial autosomal-recessive EOPD [42]. Furthermore, one locus in this gene has been identified as a genome-wide risk in European [43] and East Asian populations [44]. However, no duplications in *VPS13C* have been associated to PD so far. This gene encodes a protein that regulates lysosomal homeostasis and controls mitophagy by modulating the Pink1/Parkin pathway in cellular models [45]. Further in-depth assessment of this locus is highly warranted to tease out the potential role of duplications in this CNV segment.

Among the top CNV regions identified in our study, several regions overlapped with genes previously implicated in neurological disorders. For example, the 1q42.12 duplication includes the *DNAH14* gene, which is associated with neurodevelopmental disorders [46]. The 6q21 duplication region encompasses the *FIG4* gene, which has been linked to both CMT [47] and ALS [48]. PD typically affects different parts of the nervous system than either CMT or ALS and has a different underlying pathophysiological mechanism and clinical phenotype. Although there have been rare reported cases of individuals with PD and either CMT or ALS, suggesting a possible association, the evidence for a solid link between the two conditions is limited and remains controversial [49]. Nonetheless, the present study has expanded the genomic search to identify potential genes that could clarify the complex relationship between PD and these disorders, shedding light on partially convergent pathways leading to distinct classes of neurological disorders. The 12q24.12 duplication region contains the *ATXN2* gene, which encodes a protein with expanded trinucleotide repeats responsible for spinocerebellar ataxia type 2 (SCA2). These expanded repeats have been associated with other neurological disorders and may confer susceptibility to ALS [50] and late-onset PD [51]. Several families with *ATXN2*

medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

heterozygous expansions were previously described with predominant Parkinsonian symptoms [52,53]. However, the findings of a comprehensive analysis of polyglutamine repeat expansions conducted by the GEoPD consortium indicate that *ATXN2* does not play a significant role in the development of idiopathic Parkinson's disease [54]. The 19p13.11 duplication region harbours two genes linked to immunodeficiency: *FCHO1* and *JAK3*. Previous studies have identified shared genetic variants between PD and autoimmune/inflammatory disorders, highlighting the involvement of the immune system in the pathogenesis of PD [55]. In the prodromal phase of PD, a combination of factors including aging-related changes in the immune system and genetic/environmental influences contribute to the onset and progression of the disease, which is characterized by immunosenescence, inflammageing, and impaired adaptive immune functions [56].

Four genes within significant duplication regions showed a high degree of intolerance to loss of function and gene dosage. These genes include *ATXN2*, which is associated with susceptibility to PD [51], and *UNC13A*, a genome-wide risk locus for ALS and frontotemporal dementia [57]. *UNC13A*, identified as a potential contributor to PD, interacts with *RAB3A* to prevent alpha-synuclein-induced dopaminergic neuron degeneration [58,59].

The genes overlapping with PD-associated duplications are enriched in many pathways related to lipid (phospholipases) and steroid (estrogens) enzymatic activity and metabolism. Phospholipases have been implicated in the pathogenesis of PD through their involvement in lipid metabolism, cell signalling, and inflammatory processes [60]. Mutations in *PLA2G6*, encoding a form of phospholipase A2 enzyme, cause autosomal recessive neuroaxonal dystrophy and early-onset parkinsonism [61]. Estrogen has been shown to reduce neuroinflammation by inhibiting the NLRP3 inflammasome, thereby exerting a beneficial effect on Parkinson's disease [62]. Moreover, estrogen upregulates the expression of an anti-oxidative stress factor through estrogen receptors. Exercise has been shown to activate the anti-oxidative signalling pathway, which reduces oxidative stress and improves symptoms associated with PD, by directly activating estrogen receptors. Estrogen levels are particularly reduced in menopausal women, which increases the risk of PD due to the compromised protective effect of estrogens on dopaminergic neurons. Additionally, estrogens maintain normal mitochondrial function and dopaminergic neuron activity through the activation of estrogen-related receptors [63].

It has been established that the AAO of PD plays a pivotal role in the observed clinical heterogeneity among patients [7,64,65]. Our study shed new light on the role of CNV on the AAO with the identification of four loci with genome-wide significance for EOPD, including one deletion and three duplication regions. Of note, the three duplications were also identified in our PD GWAS (1q42.12, 6q21 and 11q12.3), while the deletion spanned the EOPD-related *PRKN* gene.

This finding aligns with the results of our burden analysis. *PRKN* homozygous or compound heterozygous deletions and duplications were identified as the most prevalent cause of EOPD and familial forms of PD [66–68]. As compared to our CNV-based GWAS analyses, we did not identify any significant differences in the overall genome-wide CNV burden, similar to the previously published

medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

study [22]. An increased burden of CNVs overlapping PD-related genes was observed in PD patients, primarily attributed to CNVs within the *PRKN* gene. This burden was even more pronounced in EOPD patients. Our observation is aligned with the published literature, in which it was shown that deletions and duplications accounted for 43.6% of all variants [14]. The presence of these variants was found to be significantly associated with an earlier onset of PD. Furthermore, PD patients with a CNV in PD-related genes manifested a significantly earlier onset of symptoms compared to those with CNVs in other genes or non-carriers.

The majority of CNVs identified in the *PRKN* gene were heterozygous, but the presence of another pathogenic variant on the other allele cannot be ruled out. Heterozygous loss of *PRKN* function may elevate the risk of developing PD [69–71] and lead to an earlier AAO [72]. However, the impact of heterozygous *PRKN* CNVs on PD susceptibility remains controversial. Some studies have indicated that carrying a heterozygous *PRKN* CNV may contribute to the development of PD, potentially through haploinsufficiency [70]. Nevertheless, a study in subjects of European ancestry demonstrated no association between heterozygous *PRKN* CNVs and EOPD [73].

Power estimation for CNV-based studies remains an interesting area to explore for future work, with the identification of CNVs in *bona fide* PD loci highlighting the robustness of our cohort to delineate CNV signals. However, certain caveats need to be considered. First, CNV identification was based on best practices applied to large-scale genotyping data, with breakpoint estimation relying on genotyped SNPs. Consequently, the resolution of the genotyping platform may have limited the accuracy of these estimates. One limitation of our study is that not all CNVs were validated using techniques such as MLPA or qPCR. Nonetheless, validation of CNVs in PD patients carrying *PRKN* and *SNCA* in one study showed a high confirmation rate, indicating the accuracy of our CNV analysis pipeline. To mitigate resolution issues, we applied a sliding window approach, nominating loci known to harbour CNVs. However, unlike traditional SNP-based GWASs, CNV-GWASs face significant challenges in identifying the causal gene(s) affected by CNVs. In our study, we identified 14 CNV regions spanning a total of 70 genes, ranging from 210 to 560 kb, exclusively in individuals of European descent. Therefore, in-depth future studies are warranted to elucidate the role of these regions in PD pathogenesis among Europeans and to replicate these findings in ethnically diverse populations.

To our knowledge, this study was the largest CNV-based GWAS performed in PD to date. Our study was able to nominate novel CNV regions for PD and confirmed that the CNV burden in EOPD was primarily driven by the *PRKN* gene. The loci nominated herein require further in-depth evaluation to define the role of these loci in the aetiopathogenesis of PD.

Acknowledgments

This study used data from the Courage-PD consortium, conducted under a partnership agreement among 35 studies. The Courage-PD consortium is supported by the EU Joint Program for Neurodegenerative Disease research (JPND https://neurodegenerationresearch.eu). This work was

medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

supported by the de.NBI Cloud within the German Network for Bioinformatics Infrastructure (de.NBI) and ELIXIR-DE in particular the de.NBI Cloud Tübingen. P. May was funded by the Fonds National de Recherche (FNR), Luxembourg, as part of the National Centre of Excellence in Research on Parkinson's Disease (NCER-PD, FNR11264123). Z. Landoulsi and P. May were supported by the DFG Research Unit FOR2715 (INTER/DFG/17/ 11583046), FOR2488 (INTER/DFG/19/14429377) and the National Centre for Excellence in Research on Parkinson's disease (NCER-PD). A.B. Singleton, D.G. Hernandez, and C. Edsall are funded by the Intramural Research Program of the National Institute on Aging, National Institutes of Health, Department of Health and Human Services, project ZO1 AG000949. E. Rogaeva is funded by the Canadian Consortium on Neurodegeneration in Aging. S.Koks is funded by MSWA. P. Taba is the recipient of an Estonian Research Council Grant PRG957. E.M.Valente is funded by the Italian Ministry of Health (Ricerca Corrente 2021). S. Bardien and J. Carr are supported by grants from the National Research Foundation of South Africa (grant number: 106052); the South African Medical Research Council (Self-Initiated Research Grant); and Stellenbosch University, South Africa; they also acknowledge the support of the NRF-DST Centre of Excellence for Biomedical Tuberculosis Research; South African Medical Research Council Centre for Tuberculosis Research; and Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town. P. Pastor have received funding from the Spanish Ministry of Science and Innovation (SAF2013-47939-R). K. Wirdefeldt and N.L. Pedersen are funded by the Swedish Research Council, grant numbers K2002-27X-14056-02B, 521-2010-2479, 521-2013-2488, and 2017-02175. N.L. Pedersen is funded by the National Institutes of Health, grant numbers ES10758 and AG 08724. C. Ran is funded by the Märta Lundkvist Foundation, Swedish Brain Foundation, and Karolinska Institutet Research Fund. A.C. Belin is funded by the Swedish Brain Foundation, Swedish Research Council, and Karolinska Institutet Research Funds. M. Tan is funded by the Parkinson's UK. M. Sharma was supported by grants from the German Research Council (DFG/SH 599/6-1), MSA Coalition, and The Michael J. Fox Foundation (USA Genetic Diversity in PD Program: GAP-India Grant ID: 17473). A. Elbaz reports grants from Agence nationale de recherche (ANR), The Michael J. Fox Foundation, Plan Ecophyto (French Ministry of Agriculture), and France Parkinson outside the submitted work. PG GEN sample collection was funded by the MRC and UK Medical Research Council (C.E. Clarke and K.E. Morrison). The sponsors had no role in the study design, data collection, data analysis, data interpretation, writing of the report, or decision to submit the paper for publication.

Author contributions

ZL, PM, MS, RK contributed to the conception and design of the study.

Z.L, A.A.K.S, C.S, D.R.B, L.M, C.L, L.M.N, E.H, C.D, L.Pavelka, P-E.S, M.R, P.L, B.P, C.E, J.K, D.G.H, C.B, G.D.M, A.Zimprich, W.P, M.Tan, E.R, A.L, S.K, P.T, S.L, A.B, J-C.C, M-C.C-H, E.M,

medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

K.B, A.B.D, G.M.H, E.D, L.Stefanis, A.M.S, E.M.V, S.P, L.Straniero, A.Zecchinelli, G.P, L.B, C.F, G.A, A.Q, M.G, L.F.B, H.M, A.N, N.H, K.N, S.J.C, Y.J.K, P.K, B.P.C.V.D.W, B.R.B, A.B.S, M.Toft, L.Pihlstrom, L.C.G, J.J.F, S.B, J.C, E.T, M.E, P.P, K.W, N.L.P, C.R, A.C.B, A.P, C.E.C, K.E.M, D.K, M.J.F, A.E, T.G, D.L contributed to the acquisition and analysis of data.

ZL, PM, MS, RK contributed to drafting the text or preparing the figures.

All the authors contributed to revise the manuscript and approved the submitted version.

Data and code availability

Individual-level CNVs and genotyping data are available on request from the COURAGE-PD consortium. Relevant scripts used in the present work are available on GitHub (https://gitlab.lcsb.uni.lu/genomeanalysis/cnv_gwas_courage-pd). For the sliding window association method, we used the code available under the Talkowski Laboratory (Massachusetts General Hospital & The Broad Institute) repository in Zenodo (https://github.com/talkowski-lab/rCNV2/tree/v1.0, with https://zenodo.org/records/6647918).

medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

References

1. Steinmetz JD, Seeher KM, Schiess N, et al. Global, regional, and national burden of disorders affecting the nervous system, 1990–2021: a systematic analysis for the Global Burden of Disease Study 2021. The Lancet Neurology 2024;23(4):344–381.

2. Lim S-Y, Klein C. Parkinson's Disease is Predominantly a Genetic Disease. JPD 2024;14(3):467–482.

3. Dorsey ER, Bloem BR. Parkinson's Disease Is Predominantly an Environmental Disease. JPD 2024;14(3):451–465.

4. Kasten M, Hartmann C, Hampf J, et al. Genotype□Phenotype Relations for the Parkinson's Disease Genes *Parkin* , *PINK1* , *DJ1:* MDSGene Systematic Review. Movement Disorders 2018;33(5):730–741.

5. Blauwendraat C, Nalls MA, Singleton AB. The genetic architecture of Parkinson's disease. The Lancet Neurology 2020;19(2):170–178.

6. Nalls MA, Blauwendraat C, Vallerga CL, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. The Lancet Neurology 2019;18(12):1091–1102.

7. Grover S, Kumar Sreelatha AA, Pihlstrom L, et al. Genome-wide Association and Meta-analysis of Age at Onset in Parkinson Disease: Evidence From the COURAGE-PD Consortium [Internet]. Neurology 2022;99(7)[cited 2024 Apr 19] Available from: https://www.neurology.org/doi/10.1212/WNL.0000000000200699

8. Toft M, Ross OA. Copy number variation in Parkinson's disease. Genome Med 2010;2(9):62.

9. La Cognata V, D'Agata V, Cavalcanti F, Cavallaro S. Genetics of Parkinson's Disease: The Role of Copy Number Variations [Internet]. In: Dorszewska J, Kozubski W, editors. Challenges in Parkinson's Disease. InTech; 2016[cited 2024 Apr 2] Available from: http://www.intechopen.com/books/challenges-in-parkinson-s-disease/genetics-of-parkinson-s-disease-the-role-of-copy-number-variations

10. La Cognata V, Morello G, D'Agata V, Cavallaro S. Copy number variability in Parkinson's disease: assembling the puzzle through a systems biology approach. Hum Genet 2017;136(1):13–37.

11. Miano-Burkhardt A, Alvarez Jerez P, Daida K, et al. The Role of Structural Variants in the Genetic Architecture of Parkinson's Disease. IJMS 2024;25(9):4801.

12. Singleton AB, Farrer M, Johnson J, et al. α-Synuclein Locus Triplication Causes Parkinson's Disease. Science 2003;302(5646):841–841.

13. Ahn T-B, Kim SY, Kim JY, et al. α-Synuclein gene duplication is present in sporadic Parkinson disease. Neurology 2008;70(1):43–49.

14. Menon PJ, Sambin S, Criniere-Boizet B, et al. Genotype–phenotype correlation in PRKN-associated Parkinson's disease. npj Parkinsons Dis. 2024;10(1):72.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

15. Pankratz N, Kissell DK, Pauciulo MW, et al. *Parkin* dosage mutations have greater pathogenicity in familial PD than simple sequence mutations. Neurology 2009;73(4):279–286.

16. Kay DM, Stevens CF, Hamza TH, et al. A comprehensive analysis of deletions, multiplications, and copy number variations in PARK2. Neurology 2010;75(13):1189–1194.

17. Ambroziak W, Koziorowski D, Duszyc K, et al. Genomic instability in the PARK2 locus is associated with Parkinson's disease. J Appl Genetics 2015;56(4):451–461.

18. Cazeneuve C, Sân C, Ibrahim SA, et al. A new complex homozygous large rearrangement of the PINK1 gene in a Sudanese family with early onset Parkinson's disease. Neurogenetics 2009;10(3):265–270.

19. Darvish H, Movafagh A, Omrani MD, et al. Detection of copy number changes in genes associated with Parkinson's disease in Iranian patients. Neuroscience Letters 2013;551:75–78.

20. Bonifati V, Rizzu P, Van Baren MJ, et al. Mutations in the *DJ-1* Gene Associated with Autosomal Recessive Early-Onset Parkinsonism. Science 2003;299(5604):256–259.

21. Macedo MG, Verbaan D, Fang Y, et al. Genotypic and phenotypic characteristics of Dutch patients with early onset Parkinson's disease. Movement Disorders 2009;24(2):196–203.

22. Sarihan EI, Pérez□Palma E, Niestroj L, et al. Genome□Wide Analysis of Copy Number Variation in Latin American Parkinson's Disease Patients. Mov Disord 2021;36(2):434–441.

23. Pankratz N, Dumitriu A, Hetrick KN, et al. Copy Number Variation in Familial Parkinson Disease. PLoS ONE 2011;6(8):e20988.

24. Grover S, Kumar□Sreelatha AA, Bobbili DR, et al. Replication of a Novel Parkinson's Locus in a European Ancestry Population. Movement Disorders 2021;36(7):1689–1695.

25. Domenighetti C, Sugier P, Ashok Kumar Sreelatha A, et al. Dairy Intake and Parkinson's Disease: A Mendelian Randomization Study. Movement Disorders 2022;37(4):857–864.

26. Mehanna R, Smilowska K, Fleisher J, et al. Age Cutoff for Early□Onset Parkinson's Disease: Recommendations from the International Parkinson and Movement Disorder Society Task Force on Early Onset Parkinson's Disease. Movement Disord Clin Pract 2022;9(7):869–878.

27. Wang K, Li M, Hadley D, et al. PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 2007;17(11):1665–1674.

28. Niestroj L-M, Perez-Palma E, Howrigan DP, et al. Epilepsy subtype-specific copy number burden observed in a genome-wide study of 17 458 subjects. Brain 2020;143(7):2106–2118.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

29. Montanucci L, Lewis-Smith D, Collins RL, et al. Genome-wide identification and phenotypic characterization of seizure-associated copy number variations in 741,075 individuals. Nat Commun 2023;14(1):4392.

30. Purcell S, Neale B, Todd-Brown K, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics 2007;81(3):559–575.

31. MacDonald JR, Ziman R, Yuen RKC, et al. The Database of Genomic Variants: a curated collection of structural variation in the human genome. Nucl. Acids Res. 2014;42(D1):D986–D992.

32. Bragin E, Chatzimichali EA, Wright CF, et al. DECIPHER: database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. Nucl. Acids Res. 2014;42(D1):D993–D1000.

33. Collins RL, Glessner JT, Porcu E, et al. A cross-disorder dosage sensitivity map of the human genome. Cell 2022;185(16):3041-3055.e25.

34. Wakefield J. Bayes factors for genome wide association studies: comparison with *P* values. Genetic Epidemiology 2009;33(1):79–86.

35. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 2020;581(7809):434–443.

36. Kuleshov MV, Jones MR, Rouillard AD, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 2016;44(W1):W90–W97.

37. Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, et al. The DisGeNET knowledge platform for disease genomics: 2019 update. Nucleic Acids Research 2019;gkz1021.

38. Therneau TM. survival: Survival Analysis [Internet]. 2001;3.7–0.[cited 2024 Jul 9] Available from: https://CRAN.R-project.org/package=survival

39. Lerche S, Wurster I, Röben B, et al. CSF NFL in a Longitudinally Assessed PD Cohort: Age Effects and Cognitive Trajectories. Movement Disorders 2020;35(7):1138–1144.

40. Ibáñez P, Bonnet A-M, Débarges B, et al. Causal relation between α-synuclein locus duplication as a cause of familial Parkinson's disease. The Lancet 2004;364(9440):1169–1171.

41. Chartier-Harlin M-C, Kachergus J, Roumier C, et al. α-synuclein locus duplication as a cause of familial Parkinson's disease. The Lancet 2004;364(9440):1167–1169.

42. Lesage S, Drouet V, Majounie E, et al. Loss of VPS13C Function in Autosomal-Recessive Parkinsonism Causes Mitochondrial Dysfunction and Increases PINK1/Parkin-Dependent Mitophagy. The American Journal of Human Genetics 2016;98(3):500–513.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

43. 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. Nat Genet 2014;46(9):989–993.

44. Foo JN, Tan LC, Irwan ID, et al. Genome-wide association study of Parkinson's disease in East Asians. Hum. Mol. Genet. 2016;ddw379.

45. Yahya V, Di Fonzo A, Monfrini E. Genetic Evidence for Endolysosomal Dysfunction in Parkinson's Disease: A Critical Overview. IJMS 2023;24(7):6338.

46. Li J, Yuan Y, Liu C, et al. *DNAH14* variants are associated with neurodevelopmental disorders. Human Mutation 2022;43(7):940–949.

47. Beloribi-Djefaflia S, Morales RJ, Fatehi F, et al. Clinical and genetic features of patients suffering from CMT4J [Internet]. J Neurol 2023;[cited 2024 Aug 5] Available from: https://link.springer.com/10.1007/s00415-023-12076-4

48. Bertolin C, Querin G, Bozzoni V, et al. New *FIG 4* gene mutations causing aggressive ALS [Internet]. Euro J of Neurology 2018;25(3)[cited 2024 Aug 5] Available from: https://onlinelibrary.wiley.com/doi/10.1111/ene.13559

49. Yamaguchi M, Omori K, Asada S, Yoshida H. Epigenetic Regulation of ALS and CMT: A Lesson from Drosophila Models. IJMS 2021;22(2):491.

50. Sproviero W, Shatunov A, Stahl D, et al. ATXN2 trinucleotide repeat length correlates with risk of ALS. Neurobiology of Aging 2017;51:178.e1-178.e9.

51. Costa RG, Conceição A, Matos CA, Nóbrega C. The polyglutamine protein ATXN2: from its molecular functions to its involvement in disease. Cell Death Dis 2024;15(6):415.

52. Furtado S, Payami H, Lockhart PJ, et al. Profile of families with parkinsonism□predominant spinocerebellar ataxia type 2 (SCA2). Movement Disorders 2004;19(6):622–629.

53. Simon-Sanchez J, Hanson M, Singleton A, et al. Analysis of SCA-2 and SCA-3 repeats in Parkinsonism: Evidence of SCA-2 expansion in a family with autosomal dominant Parkinson's disease. Neuroscience Letters 2005;382(1–2):191–194.

54. Wang L, Aasly JO, Annesi G, et al. Large-scale assessment of polyglutamine repeat expansions in Parkinson disease. Neurology 2015;85(15):1283–1292.

55. Capelle CM, Ciré S, Hedin F, et al. Early-to-mid stage idiopathic Parkinson's disease shows enhanced cytotoxicity and differentiation in CD8 T-cells in females. Nat Commun 2023;14(1):7461.

56. Tansey MG, Wallings RL, Houser MC, et al. Inflammation and immune dysfunction in Parkinson disease. Nat Rev Immunol 2022;22(11):657–673.

57. Diekstra FP, Van Deerlin VM, Van Swieten JC, et al. *C9ORF72* and *UNC13A* are shared risk loci for amyotrophic lateral sclerosis and frontotemporal dementia: A genome□wide meta□analysis. Annals of Neurology 2014;76(1):120–133.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

58. Chandrasekaran S, Bonchev D. A NETWORK VIEW ON PARKINSON'S DISEASE. Computational and Structural Biotechnology Journal 2013;7(8):e201304004.

59. Huang C-C, Chiu T-Y, Lee T-Y, et al. Soluble α-synuclein facilitates priming and fusion by releasing Ca2+ from thapsigargin-sensitive Ca2+ pool in PC12 cells. Journal of Cell Science 2018;jcs.213017.

60. Biswas S, Bagchi A. Role of phospholipases in neurodegenerative disorders [Internet]. In: Phospholipases in Physiology and Pathology. Elsevier; 2023 p. 189–199.[cited 2024 May 12] Available from: https://linkinghub.elsevier.com/retrieve/pii/B978044321800200018X

61. Lin G, Lee P-T, Chen K, et al. Phospholipase PLA2G6, a Parkinsonism-Associated Gene, Affects Vps26 and Vps35, Retromer Function, and Ceramide Levels, Similar to α-Synuclein Gain. Cell Metabolism 2018;28(4):605-618.e6.

62. Dong W, Peng Q, Liu Z, et al. Estrogen plays an important role by influencing the NLRP3 inflammasome. Biomedicine & Pharmacotherapy 2023;167:115554.

63. He W, Zhang S, Qi Z, Liu W. Unveiling the potential of estrogen: Exploring its role in neuropsychiatric disorders and exercise intervention. Pharmacological Research 2024;204:107201.

64. Blauwendraat C, Heilbron K, Vallerga CL, et al. Parkinson's disease age at onset genome□wide association study: Defining heritability, genetic loci, and α□synuclein mechanisms. Movement Disorders 2019;34(6):866–875.

65. Pavelka L, Rauschenberger A, Landoulsi Z, et al. Age at onset as stratifier in idiopathic Parkinson's disease – effect of ageing and polygenic risk score on clinical phenotypes. npj Parkinsons Dis. 2022;8(1):102.

66. Domingo A, Klein C. Genetics of Parkinson disease [Internet]. In: Handbook of Clinical Neurology. Elsevier; 2018 p. 211–227.[cited 2024 Apr 1] Available from: https://linkinghub.elsevier.com/retrieve/pii/B9780444632333000142

67. Ahmad A, Nkosi D, Iqbal MA. PARK2 Microdeletion or Duplications Have Been Implicated in Different Neurological Disorders Including Early Onset Parkinson Disease. Genes 2023;14(3):600.

68. Elfferich P, Verleun-Mooijman MC, Maat-Kievit JA, et al. Breakpoint mapping of 13 large parkin deletions/duplications reveals an exon 4 deletion and an exon 7 duplication as founder mutations. Neurogenetics 2011;12(4):263–271.

69. Castelo Rueda MP, Raftopoulou A, Gögele M, et al. Frequency of Heterozygous Parkin (PRKN) Variants and Penetrance of Parkinson's Disease Risk Markers in the Population-Based CHRIS Cohort. Front. Neurol. 2021;12:706145.

70. Klein C, Lohmann-Hedrich K, Rogaeva E, et al. Deciphering the role of heterozygous mutations in genes associated with parkinsonism. The Lancet Neurology 2007;6(7):652–662.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

71. Lubbe SJ, Bustos BI, Hu J, et al. Assessing the relationship between monoallelic *PRKN* mutations and Parkinson's risk. Human Molecular Genetics 2021;30(1):78–86.

72. Sun M, Latourelle JC, Wooten GF, et al. Influence of Heterozygosity for Parkin Mutation on Onset Age in Familial Parkinson Disease: The *Gene* PD Study. Arch Neurol 2006;63(6):826.

73. Yu E, Rudakou U, Krohn L, et al. Analysis of Heterozygous *PRKN* Variants and Copy Number Variations in Parkinson's Disease. Mov Disord 2021;36(1):178–187.

medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

figure legends

Figure 1: Genome-wide meta-analysis of PD. Miami plot of the CNV genome-wide association analyses illustrating the -log10 transformed Bonferroni-corrected p-values (DEL and DUP for deletions at the top and duplications at the bottom, mirrored respectively) of the Fisher's exact tests for the enrichment of CNVs in cases vs. controls for each 200 kb sliding window. Adjacent chromosomes are shown in alternating light and dark colors. Genomic regions that exceeded the Bonferroni-corrected significance threshold (blue line, $\alpha = 3.74\times10^{-6}$) were annotated with the genomic band containing the signal.

Figure 2: Genome-wide meta-analysis of EOPD. Miami plot of the CNV genome-wide association analyses illustrating the -log10 transformed Bonferroni-corrected p-values (DEL and DUP for deletions at the top and duplications at the bottom, mirrored respectively) of the Fisher's exact tests for the enrichment of CNVs in cases vs. controls for each 200 kb sliding window. Adjacent chromosomes are shown in alternating light and dark colors. Genomic regions that exceeded the Bonferroni-corrected significance threshold (blue line, $\alpha = 3.74\times10^{-6}$) were annotated with the genomic band containing the signal.

Figure 3: Enrichment analysis. Barplots of gene-set enrichment analysis for GO (A), KEGG (B), Reactome (C), and DisGeNet (D) using CNV- overlapping genes within the credible intervals. BP (biological process) and MF (molecular function).

Figure 4: Rare CNV burden in PD and Early-onset PD patients compared to controls for different categories. A. Logistic regression was used to calculate odds ratios (ORs) and Bonferroni-adjusted p-values for each CNV category, and were adjusted for age, sex, and the first five components of PCAs. Protein-coding gene categories were defined as all coding genes except PD-related genes. * Bonferroni adjusted p-values surpassing the multiple testing cut-off. **B-C**. Kaplan–Meier estimates of individuals (PD patients and controls in **B** and PD patients only in **C**) carrying a CNV in a PD-related gene and individuals with other or no CNVs. Probability: the probability of not having PD symptoms. Age: age at last visit for controls or age at onset for cases. Highlighting around curves indicates 95% confidence intervals.

Tables

Table 1: Genome-wide significant CNV regions in Parkinson's disease. Column 1: Cytoband localization of the CNV. Column 2: Credible interval showing the genomic coordinates (start and end position) of the credible CNVs region supported by genome-wide association signals containing the causal element/gene with 95%. Column 3: CNV types with DEL and DUP indicate deletions and duplications. Columns 4 and 5: the lowest p-values in each CNV region and corresponding odds ratio (OR) with a 95% confidence interval. Columns 6 and 7: list of genes overlapping the genome-wide significant CNVs regions and known disease genes in these regions with causal genes between brackets.

Cytoband	Credible interval	CNV type	Lowest p-value (neg log10 p)	OR [95% CI]	CNV overlapping genes	Known disease genes
1q42.12	chr1:224960000-225410000	DUP	23.85	21.18 [8.24-54.42]	*DNAH14*	Neurodevelopmental disorders (*DNAH14)*
4q22.1	chr4:90700000-90930000	DUP	5.47	7.68 [2.55-23.18]	*MMRN1, SNCA*	Parkinson's disease (*SNCA*)
6p21.2	chr6:38620000-39000000	DUP	5.8	5.98 [2.45-14.63]	*DNAH8, GLO1*	Spermatogenic failure (*DNAH8*)
6q21	chr6:109680000-110220000	DUP	18.9	13.01 [5.87-28.81]	*AK9, CD164, FIG4, MICAL1, PPIL6, SMPD2, ZBTB24*	Charcot-Marie-Tooth disease type 4J (*FIG4*), Amyotrophic lateral sclerosis type 11 (*FIG4*)
7q22.1	chr7:99080000-99560000	DUP	7.48	48.62 [2.97-796.61]	*CYP3A4, CYP3A43, CYP3A5, CYP3A7, CYP3A7-CYP3A51P, FAM200A, GJC3, OR2AE1, TMEM225B, TRIM4, ZKSCAN5, ZNF394, ZNF655, ZNF789, ZSCAN25*	-
7q33	chr7:134750000-134980000	DUP	6.41	41.75 [2.54-686.79]	*AGBL3, CYREN, STRA8, TMEM140, WDR91*	-
8q23.3	chr8:113780000-114340000	DUP	8.45	3.24 [2.11-4.99]	*CSMD3*	-
11q12.3	chr11:62910000-63350000	DUP	7.48	48.62 [2.97-796.61]	*LGALS12, PLAAT2, PLAAT3, PLAAT4, PLAAT5, SLC22A10, SLC22A24, SLC22A25, SLC22A9*	-
12q24.12	chr12:111790000-112010000	DUP	5.97	3.46 [1.98-6.04]	*ATXN2, PHETA1, SH2B3*	Spinocerebellar ataxia type 2 (*ATXN2*), Susceptibility to: Amyotrophic lateral sclerosis type 13 (*ATXN2*) and Late-onset Parkinson's disease (*ATXN2*),

medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

13q22.3	chr13:77490000-77780000	DUP	5.73	6.74 [2.53-17.98]	*ACOD1, CLN5, FBXL3, MYCBP2*	Neuronal ceroid lipofuscinosis (*CLN5*), Intellectual developmental disorder with short stature, facial anomalies, and speech defects (*FBXL3*)
15q22.2	chr15:62070000-62360000	DUP	7.95	5.03 [2.61-9.69]	*C2CD4A, VPS13C*	Parkinson's disease (*VPS13C*)
19p13.11	chr19:17580000-17960000	DUP	6.64	7.5 [2.83-19.91]	*B3GNT3, COLGALT1, FCHO1, INSL3, JAK3, MAP1S, NIBAN3, PGLS, SLC27A1, UNC13A*	Immunodeficiency 76 (*FCHO1*), T-B+ severe combined immunodeficiency (*JAK3*) Brain small vessel disease 3 (*COLGALT1*)
20p13	chr20:3380000-3740000	DUP	7.81	3.43 [2.13-5.51]	*ADAM33, ATRN, DNAAF9, ADISSP, GFRA4, HSPA12B, SIGLEC1*	-
21q22.3	chr21:46650000-46860000	DEL	6.64	3.46 [2.05-5.85]	*COL18A1, POFUT2*	Knobloch syndrome type 1 (*COL18A1*), Glaucoma, primary closed-angle (*COL18A1*)

medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Table 2: Genome-wide significant CNV regions in Early-onset Parkinson's disease.

Column 1: Cytoband localization of the CNV. Column 2: Credible interval showing the genomic coordinates (start and end position) of the credible CNVs region supported by genome-wide association signals containing the causal element/gene with 95%. Column 3: CNV types with DEL and DUP indicate deletions and duplications. Columns 4 and 5: the lowest p-values in each CNV region and corresponding odds ratio (OR) with 95% confidence interval. Columns 6 and 7: list of genes overlapping the genome-wide significant CNVs regions and known disease genes in these regions with causal genes between brackets.

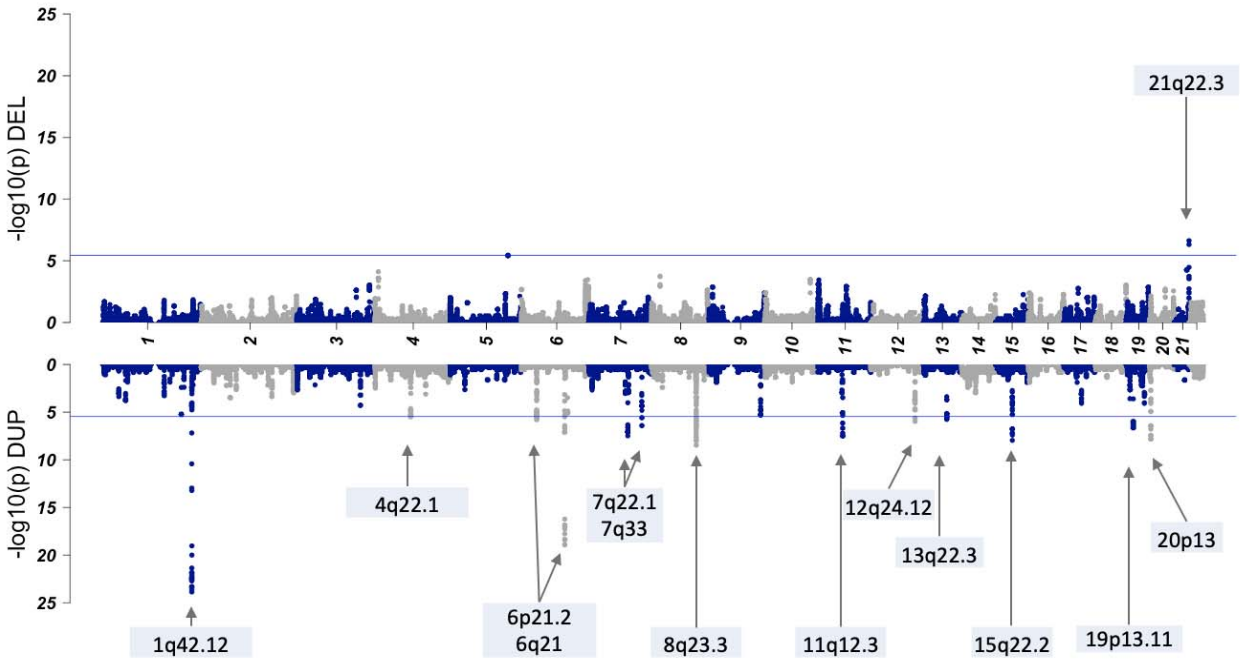
Cytoband	Credible interval	CNV type	Lowest p-value (neg log10 p)	OR [95% CI]	CNV overlapping genes	Known disease genes
1q42.12	chr1:224980000-225410000	DUP	23.85	21.18 [8.24-54.42]	*DNAH14*	Neurodevelopmental disorders (*DNAH14)*
6q21	chr6:109710000-110220000	DUP	18.9	13.01 [5.87-28.81]	*AK9, CD164, FIG4, MICAL1, PPIL6, SMPD2, ZBTB24*	Charcot-Marie-Tooth disease type 4J (*FIG4*), Amyotrophic lateral sclerosis type 11 (*FIG4*)
6q26	chr6:162340000-162930000	DEL	8.26	6.8 [3.52-13.14]	*PRKN*	Parkinson's disease (*PRKN*)
11q12.3	chr11:62920000-63340000	DUP	7.48	48.62 [2.97-796.61]	*LGALS12, PLAAT2, PLAAT4, PLAAT5, SLC22A10, SLC22A25, SLC22A9*	-

medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

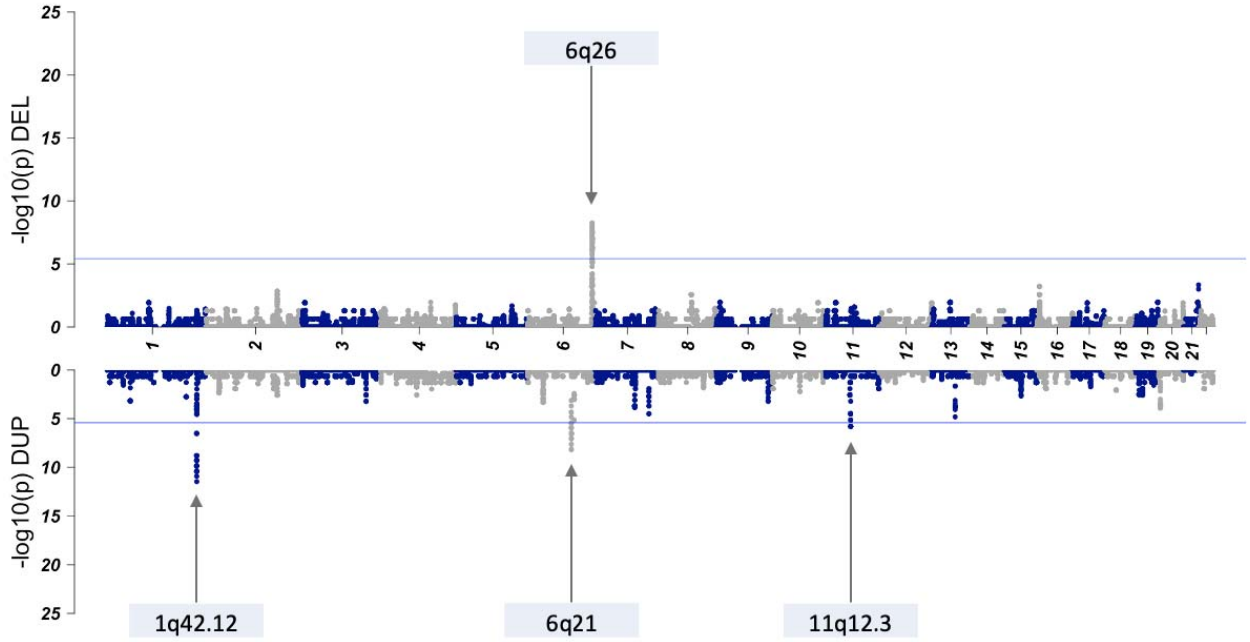
Figures:

Figure 1:



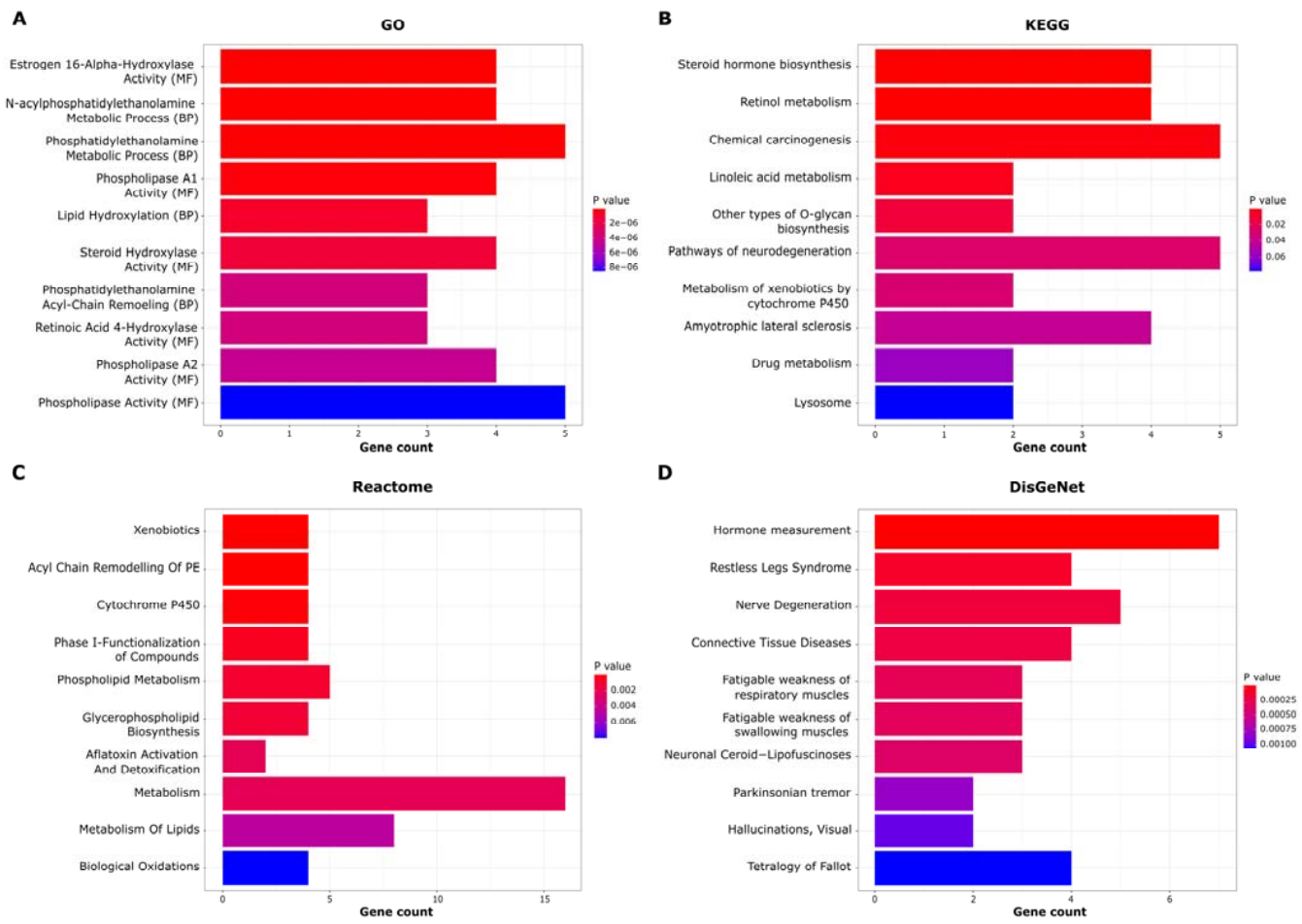
medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 2:



medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 3:

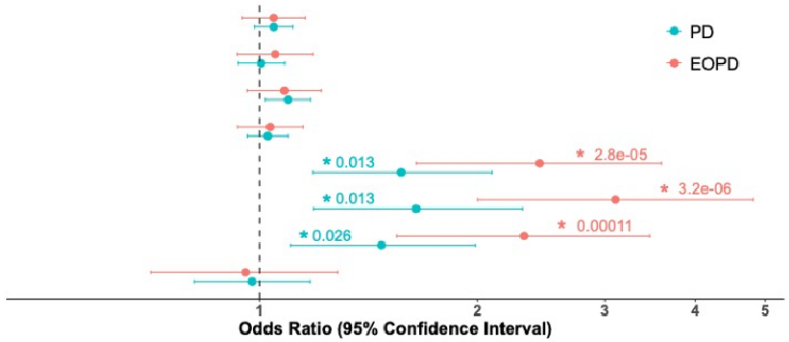


medRxiv preprint doi: https://doi.org/10.1101/2024.08.21.24311915; this version posted August 22, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

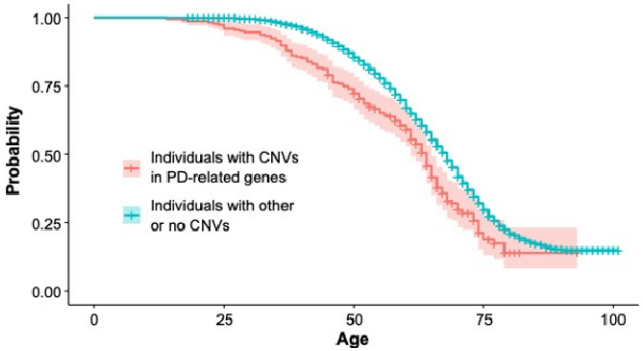
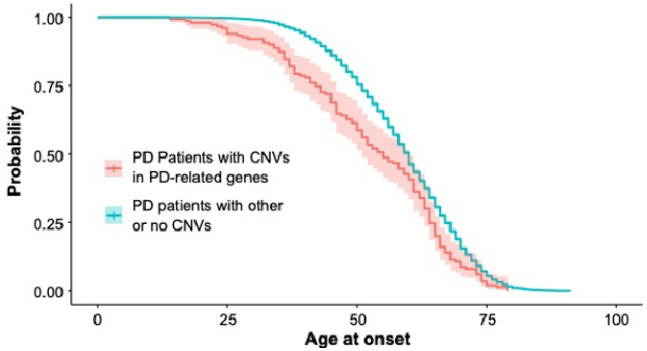
Figure 4:

A

CNV groups	Controls	PD cases	EOPD cases
All CNVs	3299/8901	3896/11035	975/2731
All Deletions	1988/8901	2090/11035	531/2731
All Duplications	1805/8901	2402/11035	592/2731
Protein-coding genes	2879/8901	3298/11035	831/2731
PD genes (exonic+intronic)	78/8901	151/11035	62/2731
PD genes (exonic)	56/8901	115/11035	53/2731
PRKN	75/8901	135/11035	57/2731
CNV > 1Mb	319/8901	283/11035	72/2731



B



C