

1 **ASSESSING THE RELATIONSHIP BETWEEN MONOALLELIC *PARK2* MUTATIONS AND PARKINSON'S** 2 **RISK**

4 **AUTHORSHIP**

5 Steven J Lubbe¹; Bernabe Bustos¹; Jing Hu¹; Dimitri Krainc¹; Theresita Joseph²; Jason Hehir³; Manuela
6 Tan²; Weijia Zhang²; Valentina Escott-Price⁴; Nigel M Williams⁴; Cornelis Blauwendraat⁵; Andrew B
7 Singleton⁵; Huw R Morris^{2*} for International Parkinson's Disease Genomics Consortium (IPDGC)⁶.

9 1 Ken and Ruth Davee Department of Neurology, and Simpson Querrey Center for Neurogenetics,
10 Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA

11 2 Department of Clinical and Movement Neurosciences, UCL Queen Square Institute of Neurology,
12 University College London, London, UK and UCL Movement Disorders Centre, UCL Queen Square
13 Institute of Neurology

14 3 National Hospital for Neurology and Neurosurgery, Queens Square, London, UK

15 4 Institute of Psychological Medicine and Clinical Neurosciences, Medical Research Council Centre
16 for Neuropsychiatric Genetics and Genomics, Department of Psychological Medicine and Neurology,
17 Cardiff University School of Medicine, Cardiff, UK

18 5 Molecular Genetics Section, Laboratory of Neurogenetics, National Institute on Aging, National
19 Institutes of Health, Bethesda, Maryland, USA

20 6 See <https://pdgenetics.org/partners>

21 **CORRESPONDING AUTHOR**

22 Professor Huw R. Morris

23 Department of Clinical and Movement Neuroscience, UCL Institute of Neurology, London, NW3 2PF,
24 United Kingdom.

25 Email: h.morris@ucl.ac.uk

26 Telephone: +44 (0)20 7830 2951

27 ABSTRACT

28 Biallelic *PARK2* (Parkin) mutations cause autosomal recessive Parkinson's (PD); however, the role of
 29 monoallelic *PARK2* mutations as a risk factor for PD remains unclear. We investigated the role of
 30 single heterozygous *PARK2* mutations in three large independent case-control cohorts totalling
 31 10,858 PD cases and 8,328 controls. Overall, after exclusion of biallelic carriers, single *PARK2*
 32 mutations were more common in PD than controls conferring a >1.5-fold increase in risk of PD
 33 ($P=0.035$), with meta-analysis (19,574 PD cases and 468,488 controls) confirming increased risk
 34 ($OR=1.65$, $P=3.69E-07$). Carriers were shown to have significantly younger ages at onset compared to
 35 non-carriers (NeuroX: 56.4 vs. 61.4 years; Exome: 38.5 vs. 43.1 years). Stratifying by mutation type,
 36 we provide preliminary evidence for a more pathogenic risk profile for single *PARK2* copy number
 37 variant (CNV) carriers compared to single nucleotide variant carriers. Studies that did not assess
 38 biallelic *PARK2* mutations or consist of predominantly early-onset cases may be biasing these
 39 estimates, and removal of these resulted in a loss of association ($OR=1.23$, $P=0.614$; $n=4$).
 40 Importantly, when we looked for additional CNVs in 30% of PD cases with apparent monoallelic
 41 *PARK2* mutations we found that 44% had biallelic mutations suggesting that previous estimates may
 42 be influenced by cryptic biallelic mutation status. While this study supports the association of single
 43 *PARK2* mutations with PD, it highlights confounding effects therefore caution is needed when
 44 interpreting current risk estimates. Together, we demonstrate that comprehensive assessment of
 45 biallelic mutation status is essential when elucidating PD risk associated with monoallelic *PARK2*
 46 mutations.

INTRODUCTION

Parkinson's (PD) is a multifactorial neurodegenerative disease. Common variation within 78 independent loci increase PD risk (1). Pathogenic mutations in autosomal dominant genes (*LRRK2*, 79 *SNCA* and *VPS35*) as well as biallelic mutations in autosomal recessive (AR) genes (*PARK2*, *DJ-1*, 80 *PINK1* and *FBXO7*) cause Mendelian PD (2). It has been suggested that single heterozygous 81 pathogenic AR mutations can increase the risk of PD, and several lines of evidence have been 82 provided for and against mutations (reviewed in Klein *et al.*, 2007) (3). Previous studies may have 83 been confounded by differences in methods for mutation detection in cases and controls. Biallelic 84 AR mutations in PD genes are rare in PD cases, but single heterozygous mutations in specific AR PD 85 genes are more common and are estimated, depending on the population, to occur in between 0.6% 86 and 3% of unaffected control individuals (4–7). Accurate estimation of any risk associated with single 87 heterozygous AR mutations is therefore essential for the counselling of biallelic carriers, monoallelic 88 carriers and their family members. Furthermore, understanding the risk associated with single AR 89 mutations may provide important insights into disease biology. Here, we investigate whether single 90 carriers of disease-causing *PARK2* mutations are at an increased risk for PD using three large 91 independent case-control cohorts using exome-focused genotype data, whole exome sequencing 92 and resequencing data from the International Parkinson's Disease Genomics Consortium (IPDGC).

RESULTS

We identified a total of 109 monoallelic *PARK2* mutation carriers in 12,251 PD cases and controls (72 93 PD, 37 controls), carrying 19 different *PARK2* variants known to cause AR PD in the biallelic state, 94 using the NeuroX genotyping platform (8). It is possible that the identified PD cases represent 95 misclassified true biallelic *PARK2* PD cases. To confirm whether PD cases carry a single pathogenic 96 allele or whether a second variant was missed, we (i) reviewed diagnostic reports if available (n=4), 97 or (ii) assessed available samples using multiplex ligation-dependent probe amplification (MLPA, 98 n=29). Of the 33 available NeuroX samples, representing ~30% of our putative monoallelic 99 individuals (5 controls, 13.5%; 28 PD-Monoallelic, 38.9%), six cases (18% of the available samples, 100 21% of available PD cases) were found to harbour a second mutation and therefore were removed, 101 leaving a total of 66 PD cases for all subsequent analyses (no controls were found to harbour a 102 second mutation).

After removal of cases with established Mendelian biallelic mutations across all known PD genes, 103 1.0% (66/6,552) of PD cases were found to harbour single heterozygous *PARK2* PD-causing 104 mutations (either heterozygous copy number variants [CNVs] or single nucleotide variants [SNVs], 105 Table 1), compared to 0.6% (37/5,693) of controls. Single heterozygote mutations might increase PD

81 risk (OR=1.55; 95% CI:1.03, 2.33; P=0.035), although ~70% of putative monogenic cases were not
 82 assessed for a second mutation. If the remaining apparent monoallelic cases had a similar rate of
 83 occult biallelic mutations, then the true underlying monoallelic carrier rate could be estimated to be
 84 lower at 0.9%, and there would be no difference between cases and controls (OR=1.39; 95% CI:0.90,
 85 2.16; P=0.117). Using age at onset (AAO) data from 5,710 (87.1%) cases, we found that NeuroX PD
 86 cases with single *PARK2* mutations have significantly lower AAOs (Average=56.4 years) than cases
 87 without known mutations (Average=61.4 years; Coeff=-5.04; 95% CI:-8.32, -1.71; P=0.003).

88 We next sought to explore the potential increased risk in two independent IPDGC case-control
 89 cohorts, using exome sequencing (cases=1,235; controls=473) (8) and resequencing data
 90 (cases=3,071; controls=2,162). We identified 28 (23 cases, 5 controls) and 52 (36 cases, 16 controls)
 91 carriers of single *PARK2* mutations in the exome (as previously described (8)) and resequencing data
 92 respectively. CNVs were not determined in the primary exome or resequencing dataset. In the
 93 exome cohort, 1.9% (23/1,231) of cases and 1.1% (5/473) of controls, and in the resequencing
 94 cohort, 1.2% (36/3,071) of cases and 0.7% (16/2,162) of controls harboured single *PARK2* SNVs.
 95 Before searching for occult second mutations, a meta-analysis of the three IPDGC (NeuroX, exome
 96 sequencing and resequencing) cohorts revealed a significant ~1.5-fold increased risk (OR=1.57; 95%
 97 CI:1.15, 2.16; P=0.005; $I^2=0.0\%$, $P_{het}=0.960$) associated with *PARK2* mutations (Figure 1). AAO data
 98 was available on 1,130 PD exome cases (91.8%) and 2,599 resequencing cases (84.6%). Albeit non-
 99 significant, exome PD cases carrying single *PARK2* SNVs had lower AAO compared non-carriers
 100 (Average=38.5 years vs. 43.1 years; Coeff=-4.34; 95% CI:-8.95, 0.28; P=0.066), with carriers having
 101 significantly lower AAO in resequencing cases (Average=52.6 years vs. 60.5 years; Coeff=-7.84; 95%
 102 CI:-12.59, -3.09; P=0.001). We then used MLPA to search for potentially missed *PARK2* CNVs in
 103 mutation carriers. Four of the nine available exome DNA samples (44%, all PD cases) were found to
 104 harbour a missed second mutation and were removed from subsequent analyses. Assuming a similar
 105 rate of occult biallelic carriage across both datasets, the true rate of monoallelic cases could be
 106 estimated to be 1.2% and 0.7% in the exome and resequencing cases as compared with 1.1% and
 107 0.7% of controls, respectively.

108 We next performed a meta-analysis of available cohorts and studies that reported heterozygous PD-
 109 causing mutation rates in cases and controls, from European ancestry cohorts only. Three cohorts
 110 (Parkinson's Progression Markers Initiative, PPMI, <https://www.ppmi-info.org/>; UK Biobank
 111 Genotyping and Exome cohorts, <https://www.ukbiobank.ac.uk/>) and 21 published studies were
 112 included in our analyses (5,6,9–27). Including our cohorts, the meta-analysis revealed a significant
 113 1.65-fold increased PD risk in single *PARK2* mutation carriers (95% CI: 1.36, 2.00; P=3.69E-07;
 114 $I^2=0.0\%$, $P_{het}=0.594$) (Figure 2).

115 As in our study, occult second mutations are likely biasing these estimates, so we therefore
116 restricted our meta-analysis to 9 studies (9/21, 43%) that searched for a second *PARK2* mutation
117 (Supplementary Table 1). Based on these studies, single *PARK2* mutations confer a 2-fold increase in
118 PD risk in carriers (OR=2.00, 95% CI:1.10, 3.62, P=0.023; I²=1.3%, P_{het}=0.423) (Supplementary Figure
119 1).

120 Inclusion of studies that used predominantly early-onset PD (EOPD) cases may additionally be
121 inflating these estimates, we therefore repeated the meta-analysis excluding these EOPD studies
122 (15/21, 71%), which demonstrated a 1.5-fold significant increased risk in carriers (OR=1.50, 95%
123 CI:1.22, 1.84; P=1.05E-04; I²=0%, P_{het}=0.927) (Supplementary Figure 2).

124 Restricting our analysis to the four non-EOPD studies that searched for biallelic carriers
125 demonstrated that single *PARK2* mutations were not associated with an increased PD risk in these
126 cohorts (OR=1.23, 95% CI:0.55, 2.75, P=0.614; I²=0.0%, P_{het}=0.657; Supplementary Figure 3).

127 The pathogenicity of the common *PARK2* p.R275W variant in AR PD is not as clear cut as other
128 *PARK2* mutations. To test whether the observed SNV association is driven by p.R275W, we repeated
129 the meta-analysis after excluding this variant. Removal of p.R275W resulted in a marginally
130 increased estimate (OR=1.76, 95% CI:1.37, 2.28, P=1.42E-05; I²=0.0%, P_{het}=0.673; Supplementary
131 Table 2; Supplementary Figure 4). Limiting our analysis to the 8 cohorts which assessed biallelic
132 mutations indicated a >2-fold increased risk (OR=2.41, 95% CI:1.17, 4.96, P=0.017; I²=0.0%, P_{het}=0.791)
133 (Supplementary Figure 5).

134 The contribution of biallelic *PARK2* CNVs to AR PD is well established; however, that of heterozygous
135 CNV carriers remains unclear. We identified monoallelic *PARK2* CNVs in 0.17% (11/6,552) of non-
136 Mendelian PD cases compared to 0.07% (4/5,693) controls (Table 1) using the NeuroX data only.
137 None of these CNV carriers overlapped with NeuroX SNV carriers. There was a >2.5-fold increase in
138 PD risk for *PARK2* CNV heterozygote carriers compared to controls (OR=2.53; 95% CI:0.80, 7.99;
139 P=0.113) but this was not statistically significant.

140 It has been suggested that monoallelic *PARK2* CNVs might confer a higher risk that is associated with
141 a more pathogenic profile compared to other AR mutations (28). To assess this, we compared
142 differences in risk between CNV and SNVs carriers in the NeuroX cohort. A total of 55 *PARK2* SNV
143 carriers were seen in non-Mendelian PD cases (55/6,552; 0.8%) compared to 33 controls (33/5,693;
144 0.6%) (OR=1.43; 95% CI:0.92, 2.21; P=0.108) (Table 1). To test whether *PARK2* CNVs confer a more
145 “pathogenic” risk profile compared to SNVs we performed AAO analysis in the NeuroX data only.
146 Carriers of heterozygous *PARK2* CNVs had a mean AAO of 58.4 years, compared to non-carriers (61.4
147 years) (Coeff=-3.11; 95% CI:-11.15, 4.93; P=0.449).

To further investigate the potential different risk profiles, we performed separate meta-analyses of published *PARK2* mutation data for SNVs and CNVs. Meta-analyses, including the current data, revealed significant independent increased PD risks for SNVs (OR=1.56, 95% CI:1.22, 2.00, P=4.46E-04; I²=0%, P_{het}=0.968) and CNVs (OR=1.85, 95% CI:1.38, 2.50, P=4.55E-05; I²=0.0%, P_{het}=0.640) (Figure 3; Supplementary Table 3). Restricting our meta-analysis to studies that searched for second hits suggested that PD risk was larger in carriers of single *PARK2* CNVs (OR=3.11, 95% CI:1.23, 7.89, P=0.016; I²=0.0%, P_{het}=0.879) compared to those harbouring heterozygous SNVs (OR=1.59, 95% CI:0.79, 3.20, P=0.191; I²=0.0%, P_{het}=0.785).

DISCUSSION

The role of rare biallelic mutations in *PARK2* in AR PD (MIM#600116) is well established. Here, using data from a large PD case-control cohort, we identified a total of 109 carriers of single heterozygous *PARK2* mutations. After exclusion of PD cases with known mutations, we demonstrated that carriers of single mutations were at a small but significantly increased risk of PD (OR=1.55; 95% CI:1.03, 2.33; P=0.035). This was confirmed by a meta-analysis with two additional IPDGC cohorts (cases=10,954; controls=8,328) which demonstrated a significant >1.5-fold increased risk (P=0.005). Carriers also had significantly lower AAOs than non-carriers (56.4 years vs. 61.4 years; P=0.003). Similar findings were seen in the exome and resequencing data for increased risk (Exome, OR=2.20; Resequencing, OR=1.59) and younger AAOs compared to non-carriers (Exome, 38.5 years vs. 43.1 years; Resequencing, 52.6 years vs. 60.5 years). A meta-analysis of 19,574 PD cases and 468,488 controls from 27 cohorts further confirmed that heterozygous *PARK2* mutations confer an increased PD risk (OR=1.65; P=3.69E-07). However, several confounding factors are likely biasing these estimates in favour of increased risk and are explored below. Large-scale studies in systematically recruited cohorts that have comprehensively interrogated biallelic *PARK2* mutations are therefore needed to accurately determine the risk associated with single mutations.

The relatively common p.R275W (c.823C>T, rs34424986) variant, the most frequent PD-associated variant in *PARK2*, has not been reported in the homozygous state and has only been reported in compound heterozygotes with another mutation in multiple AR PD families (MIM#602544), and has been classified as likely pathogenic. p.R275W reduces protein stability by disrupting binding to phosphorylated ubiquitin and results in reduced Parkin levels (29) supporting the pathogenicity of p.R275W. We examined whether the increased PD risk associated with single *PARK2* variants was driven by this variant. The observation that the OR increases after removal suggests that p.R275W may have reduced effect on enzyme activity compared to other mutations, and that, due to its more

181 common frequency, its presence may be diluting the true effect of heterozygous *PARK2* mutations in
182 PD biology.

183 Our analysis provides some support for a more “pathogenic” risk profile associated with *PARK2* CNVs
184 as compared with SNVs (CNVs, OR=2.53; SNVs, OR=1.43) in the NeuroX data. Both mutation types
185 appear to be associated with lower AAOs; however, the small number of observed CNVs prevents
186 any definitive conclusions from being drawn. While the meta-analysis results here support the
187 increased risk for *PARK2* CNV carriers (OR=1.85), albeit marginally higher than *PARK2* SNVs
188 (OR=1.56), the increased risk associated with *PARK2* CNVs should be interpreted with care. Small
189 sample sizes, low CNV frequency and failure to investigate/report CNVs may have resulted in an
190 underestimated effect size seen in the meta-analysis. Failure to conclusively look for second *PARK2*
191 hits may also be a potential confounder when trying to estimate the risk associated with single
192 *PARK2* mutations. Additional work in larger cohorts where both *PARK2* SNVs and CNVs are routinely
193 assessed is therefore needed to gain more accurate insight into the different risk profiles associated
194 with different mutation types.

195 This large study builds on previous work looking at single *PARK2* mutations in PD aetiology. While
196 several failed to identify single known pathogenic *PARK2* mutations in controls thereby supporting
197 increased disease risk, others have found equal frequencies in both cases and controls providing
198 evidence against increased risk (4,7,9,10,24,30–32). These estimates have, however, been based on
199 relatively small sample sets which have made it difficult to conclusively determine if single mutations
200 confer any risk. The inclusion of non-ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>) variants
201 represents a potential confounder in that we may be overestimating the frequency of disease
202 relevant single *PARK2* mutations. Limiting our analyses to ClinVar variants only did not result in
203 considerably different risk estimates across all comparisons (All studies, OR=1.70, P=2.65E-07;
204 Biallelic studies, OR=1.99, P=0.036; non-EOPD studies, OR=1.55, P=5.6E-05). Another confounder
205 relates to the fact that we observed a significant rate of occult second pathogenic mutations in
206 putative monoallelic cases in our NeuroX cohort. The detected rate of occult biallelic carrier status
207 was high in our two datasets (6/28, 21% and 4/9, 44%), approaching one half of PD cases with
208 apparent monoallelic status. Additionally, several studies included in the analyses here have not
209 searched for potentially hidden biallelic mutations in all cases and controls or have only interrogated
210 a subset of *PARK2* mutations. Inclusion of these PD cases in our analysis is likely to appreciably
211 influence our estimate. However, restricting the meta-analyses to 9 cohorts that searched for
212 biallelic *PARK2* mutations in all cases and controls demonstrated that single mutations confer a 2-
213 fold increase in risk in carriers. A further confounding factor is the use of EOPD cases (<50 years) in
214 such studies which may be additionally inflating risk estimates as *PARK2* mutations are more likely to

215 occur in PD cases of younger onset. This was observed in the IPDGC cohorts, with a higher estimate
216 in the exome cohort compared to the NeuroX and resequencing cohorts. The additional removal of
217 predominantly EOPD studies resulted in the loss of the original association (OR=1.23; P=0.614) but
218 this was based on a few small studies (n=4). This suggests that current estimates of the effect of
219 single mutations in modulating PD risk may not be accurate, but also stresses the importance of
220 comprehensively searching for biallelic mutations in systematically recruited cohorts. It remains
221 possible that there are further “occult” non-coding mutations affecting the promotor or splicing that
222 have not yet been identified.

223 There are some limitations to our study. NeuroX biallelic cases will have been missed as not all
224 possible PD-causing variants are represented on the chip (33). The same applies to the detection of
225 biallelic carriers in the UK Biobank genotyping cohort. Identifying *PARK2* CNVs from NeuroX SNP
226 genotype data using PennCNV may have missed smaller deletions/duplications. The false positive
227 rate of PennCNV as a method for CNV detection was estimated to be 9.0-17.7%, with false positive
228 CNVs predominantly small in size and occurring regardless of genotyping chip used (34). Our CNV
229 detection false-positive rate in the NeuroX cohort is 6.9%. However, the fact that (i) the NeuroX
230 variants are not evenly distributed across the *PARK2* locus (accounts for four misclassified samples),
231 and (ii) we were looking for CNVs as small as a single exon, may have resulted in our approach
232 missing or inaccurately calling CNVs in our large cohort comprising predominantly late-onset PD
233 cases. There are limitations in defining CNVs from IPDGC and UK Biobank exome data, so CNVs were
234 only investigated using MLPA in identified PD-Monoallelic exome cases. As the exome cohort
235 predominantly consists of EOPD cases, it is likely that additional *PARK2* CNVs carriers were
236 undetected. We therefore sought to validate the monoallelic status of available carriers by accessing
237 diagnostic reports or directly assessing CNVs using MLPA and discovered a high rate of undetected
238 second hits in both our datasets. Previous studies which have not systematically searched for second
239 hits may have therefore erroneously determined the monoallelic carrier rate meaning that the
240 estimates derived from our in-house cohorts and other published meta-analyses may not be
241 accurate. It is therefore very important that any proposed increased risk associated with single
242 *PARK2* mutations be considered with caution as, based on findings presented here, a substantial part
243 of the reported excess on monoallelic carriers may relate to occult biallelic status.

244 In conclusion, while much of the data demonstrates that harbouring a single heterozygous *PARK2*
245 mutation increases PD risk and that single *PARK2* CNVs may be more pathogenic than *PARK2* SNVs,
246 there may be confounding factors. This is supported by our finding of no increased risk associated
247 with single *PARK2* mutations upon restricting our analysis to studies that assessed biallelic mutations
248 in cases and controls, and studies that did not include predominantly EOPD cases. Before the risk

associated with single heterozygous mutations can be accurately defined, we highlight the importance of assessing ‘second hits’ in all cases and controls where both SNVs and CNVs are systematically interrogated in large-scale cohorts that have been systematically recruited.

252

253 METHODS AND MATERIALS

High-quality genotype data from the NeuroX chip on 6,558 PD cases and 5,693 controls was assessed as part of the IPDGC (dbGaP Study Accession number: phs000918.v1.p1). Sample collection and variant genotyping have been described elsewhere (33). IPDGC exome sequencing data from 1,235 PD cases and 473 controls were used as a replication cohort (EGA Study Accession numbers: EGAS00001002103, EGAS00001002110, EGAS00001002113, EGAS00001002156; dbGaP Study Accession number: phs001103.v1.p1), and is described elsewhere (8). Additional replication cohorts used include: IPDGC resequencing cohort (cases=3,071, controls=2,162), UK Biobank Genotyping (cases=1,428, controls=312,098; downloaded April 2018, under application number 33601) and UK Biobank exome sequencing (cases=114, controls=38,263; downloaded May 2019) cohorts and the PPMI Exome sequencing *de novo* cohort (cases=385, controls=179). Duplicate samples were removed, where possible, from all analyses. Samples with missing call rates >5% were excluded during quality control. Variants (excluding synonymous) from known Mendelian PD-causing genes were extracted. Pathogenic mutations were identified as previously described (8). Rare *PARK2* (NM_013988 and NM_004562) CNVs were identified in the NeuroX cohort using PennCNV (34). CNVs spanning a minimum of ten variants were selected and visually confirmed. Monoallelic PD cases were defined as those carrying a single heterozygous pathogenic *PARK2* allele as defined according to OMIM (<http://omim.org/>), the Movement Disorder Society Genetic mutation database (<https://www.mdsgene.org/>) or the Parkinson Disease Mutation Database (<http://www.molgen.vib-ua.be/PDMutDB/>) (Supplementary Table 4). Where available, samples were investigated by (i) accessing sample diagnostic records, or (ii) using MLPA (SALSA P051 v.D1 probe mix [MRC-Holland, The Netherlands]) to confirm their monoallelic status. Without phasing information, any two PD-causing hits identified in an individual are assumed to be in trans.

To assess whether PD risk might be associated with (i) all monoallelic variants, (ii) CNVs alone or (iii) SNVs alone, as indicated by case-control differences, we used logistic regression correcting for gender and principal components (C1-4). Linear regression was used to investigate the impact of single AR mutations on AAO.

A literature review was undertaken (on 01/10/2019) to identify published data on heterozygous *PARK2* mutations, using search terms including combinations of the following terms: Parkinson’s disease, PD, *Parkin*, *PARK2* and heterozygous. Additional studies were identified by manual search of

283 references cited in published articles. Should any of the studies include previously published data,
284 the most recent data was selected where possible. Meta-analysis was conducted using standard
285 methods modelling fixed effects, using Cochran's Q-statistic to test for heterogeneity (P_{het}) (35) and
286 the I^2 statistic (36) to quantify the proportion of the total variation caused by heterogeneity relating
287 to possible differences in sample recruitment and assessment between studies. Meta-analyses were
288 performed for CNVs and SNVs separately to investigate potential different risk profiles for each
289 mutation type.

290

291 **ACKNOWLEDGEMENTS**

292 We would like to thank all of the subjects who donated their time and biological samples to be a part
293 of this study. We also would like to thank all members of the International Parkinson Disease
294 Genomics Consortium (IPDGC). See for a complete overview of members, acknowledgements and
295 funding <http://pdgenetics.org/partners>.

296 This study was supported by Parkinson's UK (grants 8047 and J-0804) and the Medical Research
297 Council (G0700943 and G1100643). This work was supported in part by the Intramural Research
298 Programs of the National Institute of Neurological Disorders and Stroke (NINDS), the National
299 Institute on Aging (NIA), and the National Institute of Environmental Health Sciences both part of the
300 National Institutes of Health, Department of Health and Human Services; project numbers 1ZIA-
301 NS003154, Z01-AG000949-02 and Z01-ES101986. This research has been conducted using the UK
302 Biobank Resource. In addition, this work was supported by the Department of Defense (award
303 W81XWH-09-2-0128), and The Michael J Fox Foundation for Parkinson's Research. This work was
304 supported by National Institutes of Health grants R01NS037167, R01CA141668, P50NS071674,
305 American Parkinson Disease Association (APDA); Barnes Jewish Hospital Foundation; Greater St Louis
306 Chapter of the APDA. The KORA (Cooperative Research in the Region of Augsburg) research platform
307 was started and financed by the Forschungszentrum für Umwelt und Gesundheit, which is funded by
308 the German Federal Ministry of Education, Science, Research, and Technology and by the State of
309 Bavaria. This study was also funded by the German Federal Ministry of Education and Research
310 (BMBF) under the funding code 031A430A, the EU Joint Programme - Neurodegenerative Diseases
311 Research (JPND) project under the aegis of JPND -www.jpnd.eu- through Germany, BMBF, funding
312 code 01ED1406 and iMed - the Helmholtz Initiative on Personalized Medicine. This study is funded
313 by the German National Foundation grant (DFG SH599/6-1) (grant to M.S), Michael J Fox Foundation,
314 and MSA Coalition, USA (to M.S). The French GWAS work was supported by the French National
315 Agency of Research (ANR-08-MNP-012). This study was also funded by France-Parkinson Association,
316 Fondation de France, the French program "Investissements d'avenir" funding (ANR-10-IAIHU-06) and

a grant from Assistance Publique-Hôpitaux de Paris (PHRC, AOR-08010) for the French clinical data. This study was also sponsored by the Landspítali University Hospital Research Fund (grant to SSv); Icelandic Research Council (grant to SSv); and European Community Framework Programme 7, People Programme, and IAPP on novel genetic and phenotypic markers of Parkinson's disease and Essential Tremor (MarkMD), contract number PIAP-GA-2008-230596 MarkMD (to HP and JHu). Institutional research funding IUT20-46 was received of the Estonian Ministry of Education and Research (SK). The McGill study was funded by the Michael J. Fox Foundation and the Canadian Consortium on Neurodegeneration in Aging (CCNA). This study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, Md. (<http://biowulf.nih.gov>), and DNA panels, samples, and clinical data from the National Institute of Neurological Disorders and Stroke Human Genetics Resource Center DNA and Cell Line Repository. People who contributed samples are acknowledged in descriptions of every panel on the repository website. We thank the French Parkinson's Disease Genetics Study Group and the Drug Interaction with genes (DIGPD) study group: Y Agid, M Anheim, F Artaud, A-M Bonnet, C Bonnet, F Bourdain, J-P Brandel, C Brefel-Courbon, M Borg, A Brice, E Broussolle, F Cormier-Dequaire, J-C Corvol, P Damier, B Debilly, B Degos, P Derkinderen, A Destée, A Dürr, F Durif, A Elbaz, D Grabli, A Hartmann, S Klebe, P. Krack, J Kraemmer, S Leder, S Lesage, R Levy, E Lohmann, L Lacomblez, G Mangone, L-L Mariani, A-R Marques, M Martinez, V Mesnage, J Muellner, F Ory-Magne, F Pico, V Planté-Bordeneuve, P Pollak, O Rascol, K Tahiri, F Tison, C Tranchant, E Roze, M Tir, M Vérin, F Viallet, M Vidailhet, A You. We also thank the members of the French 3C Consortium: A Alépérovitch, C Berr, C Tzourio, and P Amouyel for allowing us to use part of the 3C cohort, and D Zelenika for support in generating the genome-wide molecular data. We thank P Tienari (Molecular Neurology Programme, Biomedicum, University of Helsinki), T Peuralinna (Department of Neurology, Helsinki University Central Hospital), L Myllykangas (Folkhalsan Institute of Genetics and Department of Pathology, University of Helsinki), and R Sulkava (Department of Public Health and General Practice Division of Geriatrics, University of Eastern Finland) for the Finnish controls (Vantaa85+ GWAS data). We used genome-wide association data generated by the Wellcome Trust Case-Control Consortium 2 (WTCCC2) from UK patients with Parkinson's disease and UK control individuals from the 1958 Birth Cohort and National Blood Service. Genotyping of UK replication cases on ImmunoChip was part of the WTCCC2 project, which was funded by the Wellcome Trust (083948/Z/07/Z). UK population control data was made available through WTCCC1. This study was supported by the Medical Research Council and Wellcome Trust disease centre (grant WT089698/Z/09/Z to NW, JHa, and AS). As with previous IPDGC efforts, this study makes use of data generated by the Wellcome Trust Case-Control Consortium. A full list of the investigators who contributed to the generation of

the data is available from www.wtccc.org.uk. Funding for the project was provided by the Wellcome Trust under award 076113, 085475 and 090355. Sequencing and genotyping done in McGill University was supported by grants from the Michael J. Fox Foundation, the Canadian Consortium on Neurodegeneration in Aging (CCNA), the Canada First Research Excellence Fund (CFREF), awarded to McGill University for the Healthy Brains for Healthy Lives (HBHL) program and Parkinson's Society Canada. We thank Jeffrey Barrett and Jason Downing (Illumina Inc) for assistance with the design of the ImmunoChip and NeuroX arrays. DNA extraction work that was done in the UK was undertaken at University College London Hospitals, University College London, who received a proportion of funding from the Department of Health's National Institute for Health Research Biomedical Research Centres funding. This study was supported in part by the Wellcome Trust/Medical Research Council Joint Call in Neurodegeneration award (WT089698) to the Parkinson's Disease Consortium (UKPDC), whose members are from the UCL Institute of Neurology, University of Sheffield, and the Medical Research Council Protein Phosphorylation Unit at the University of Dundee. We thank the Quebec Parkinson's Network (<http://rpq-qpn.org>) and its members. This work was supported by the Medical Research Council grant MR/N026004/1. The Braineac project was supported by the MRC through the MRC Sudden Death Brain Bank Grant (MR/G0901254) to J.H. P.A.L. was supported by the MRC (grants MR/N026004/1 and MR/L010933/1) and Michael J. Fox Foundation for Parkinson's Research. This work was supported by the Simpson Querrey Center for Neurogenetics (to DK).

369

370 CONFLICT OF INTEREST STATEMENT

371

372 H.R.M reports grants from Medical Research Council UK, grants from Wellcome Trust, grants from Parkinson's UK, grants from Ipsen Fund, during the conduct of the study; grants from Motor Neuron Disease Association, grants from Welsh Assembly Government, personal fees from Teva, personal fees from Abbvie, personal fees from Teva, personal fees from UCB, personal fees from Boehringer-Ingelheim, personal fees from GSK, outside the submitted work. D.K. is the Founder and Scientific Advisory Board Chair of Lysosomal Therapeutics Inc. D.K. serves on the scientific advisory boards of The Silverstein Foundation, Intellia Therapeutics, and Prevail Therapeutics and is a Venture Partner at OrbiMed.

380

381 REFERENCES

- 382 1. Nalls, M. A., Blauwendraat, C., Vallerger, C. L., et al. (2019) Identification of novel risk loci,
383 causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide
384 association studies. *Lancet Neurol.*, **18**, 1091–1102.

- 385 2. Lubbe, S. and Morris, H. R. (2014) Recent advances in Parkinson's disease genetics. *J. Neurol.*,
386 **261**, 259–66.
- 387 3. Klein, C., Lohmann-Hedrich, K., Rogaeva, E., et al. (2007) Deciphering the role of heterozygous
388 mutations in genes associated with parkinsonism. *Lancet Neurol.*, **6**, 652–662.
- 389 4. Kay, D. M., Moran, D., Moses, L., et al. (2007) Heterozygous parkin point mutations are as
390 common in control subjects as in Parkinson's patients. *Ann. Neurol.*, **61**, 47–54.
- 391 5. Kay, D. M., Stevens, C. F., Hamza, T. H., et al. (2010) A comprehensive analysis of deletions,
392 multiplications, and copy number variations in PARK2. *Neurology*, **75**, 1189–1194.
- 393 6. Lincoln, S. J., Maraganore, D. M., Lesnick, T. G., et al. (2003) Parkin variants in North American
394 Parkinson's disease: Cases and controls. *Mov. Disord.*, **18**, 1306–1311.
- 395 7. Brüggemann, N., Mitterer, M., Lanthaler, A. J., et al. (2009) Frequency of heterozygous Parkin
396 mutations in healthy subjects: Need for careful prospective follow-up examination of
397 mutation carriers. *Park. Relat. Disord.*, **15**, 425–429.
- 398 8. Lubbe, S. J., Escott-Price, V., Gibbs, J. R., et al. (2016) Additional rare variant analysis in
399 Parkinson's disease cases with and without known pathogenic mutations: evidence for
400 oligogenic inheritance. *Hum. Mol. Genet.*, **25**, 5483–5489.
- 401 9. Kann, M., Jacobs, H., Mohrmann, K., et al. (2002) Role of parkin mutations in 111 community-
402 based patients with early-onset parkinsonism. *Ann. Neurol.*, **51**, 621–625.
- 403 10. Klein, C., Djarmati, A., Hedrich, K., et al. (2005) PINK1, Parkin, and DJ-1 mutations in Italian
404 patients with early-onset parkinsonism. *Eur. J. Hum. Genet.*, **13**, 1086–93.
- 405 11. Lesage, S., Lohmann, E., Tison, F., et al. (2008) Rare heterozygous parkin variants in French
406 early-onset Parkinson disease patients and controls. *J. Med. Genet.*, **45**, 43–46.
- 407 12. Macedo, M. G., Verbaan, D., Fang, Y., et al. (2009) Genotypic and phenotypic characteristics
408 of Dutch patients with early onset Parkinson's disease. *Mov. Disord.*, **24**, 196–203.
- 409 13. Moura, K. C. V., Campos Junior, M., De Rosso, A. L. Z., et al. (2013) Genetic analysis of PARK2
410 and PINK1 genes in Brazilian patients with early-onset Parkinson's disease. *Dis. Markers*, **35**,
411 181–185.
- 412 14. Pankratz, N., Dumitriu, A., Hetrick, K. N., et al. (2011) Copy number variation in familial
413 Parkinson disease. *PLoS One*, **6**, e20988.
- 414 15. Schlitter, A. M., Kurz, M., Larsen, J. P., et al. (2006) Parkin gene variations in late-onset
415 Parkinson's disease: Comparison between Norwegian and German cohorts. *Acta Neurol.*

- 416 *Scand.*, **113**, 9–13.
- 417 16. Simon-Sanchez, J., Scholz, S., Matarin, M. del M., et al. (2008) Genomewide SNP assay reveals
418 mutations underlying Parkinson disease. *Hum. Mutat.*, **29**, 315–22.
- 419 17. Sironi, F., Primignani, P., Zini, M., et al. (2008) Parkin analysis in early onset Parkinson’s
420 disease. *Parkinsonism Relat. Disord.*, **14**, 326–333.
- 421 18. Spataro, N., Roca-Umbert, A., Cervera-Carles, L., et al. (2017) Detection of genomic
422 rearrangements from targeted resequencing data in Parkinson’s disease patients. *Mov.*
423 *Disord.*, **32**, 165–169.
- 424 19. Bandrés-Ciga, S., Mencacci, N. E., Durán, R., et al. (2015) Analysis of the genetic variability in
425 Parkinson’s disease from Southern Spain. *Neurobiol. Aging*, **37**, 210.e1-210.e5.
- 426 20. Wiley, J., Lynch, T., Lincoln, S., et al. (2004) Parkinson’s disease in Ireland: Clinical
427 presentation and genetic heterogeneity in patients with parkin mutations. *Mov. Disord.*, **19**,
428 677–681.
- 429 21. Benitez, B. A., Davis, A. A., Jin, S. C., et al. (2016) Resequencing analysis of five Mendelian
430 genes and the top genes from genome-wide association studies in Parkinson’s Disease. *Mol.*
431 *Neurodegener.*, **11**, 29.
- 432 22. Bras, J., Guerreiro, R., Ribeiro, M., et al. (2008) Analysis of Parkinson disease patients from
433 Portugal for mutations in SNCA, PRKN, PINK1 and LRRK2. *BMC Neurol.*, **8**, 1.
- 434 23. Brooks, J., Ding, J., Simon-Sanchez, J., et al. (2009) Parkin and PINK1 mutations in early-onset
435 Parkinson’s disease: comprehensive screening in publicly available cases and control. *J. Med.*
436 *Genet.*, **46**, 375–381.
- 437 24. Clark, L. N., Afridi, S., Karlins, E., et al. (2006) Case-control study of the parkin gene in early-
438 onset Parkinson disease. *Arch. Neurol.*, **63**, 548–52.
- 439 25. Erer, S., Egeli, U., Zarifoglu, M., et al. (2016) Mutation analysis of the PARKIN, PINK1, DJ1, and
440 SNCA genes in Turkish early-onset Parkinson’s patients and genotype-phenotype correlations.
441 *Clin. Neurol. Neurosurg.*, **148**, 147–53.
- 442 26. Hertz, J. M., Ostergaard, K., Juncker, I., et al. (2006) Low frequency of Parkin, Tyrosine
443 Hydroxylase, and GTP Cyclohydrolase I gene mutations in a Danish population of early-onset
444 Parkinson’s Disease. *Eur. J. Neurol.*, **13**, 385–90.
- 445 27. Huttenlocher, J., Stefansson, H., Steinberg, S., et al. (2015) Heterozygote carriers for CNVs in
446 PARK2 are at increased risk of Parkinson’s disease. *Hum. Mol. Genet.*, **24**, 5637–5643.

- 447 28. Pankratz, N., Kissell, D. K., Pauciulo, M. W., et al. (2009) Parkin dosage mutations have greater
448 pathogenicity in familial PD than simple sequence mutations. *Neurology*, **73**, 279–286.
- 449 29. Yi, W., MacDougall, E. J., Tang, M. Y., et al. (2019) The landscape of Parkin variants reveals
450 pathogenic mechanisms and therapeutic targets in Parkinson’s disease. *Hum. Mol. Genet.*, **28**,
451 2811–2825.
- 452 30. Hedrich, K., Marder, K., Harris, J., et al. (2002) Evaluation of 50 probands with early-onset
453 Parkinson’s disease for Parkin mutations. *Neurology*, **58**, 1239–1246.
- 454 31. West, A., Periquet, M., Lincoln, S., et al. (2002) Complex relationship between Parkin
455 mutations and Parkinson disease. *Am. J. Med. Genet. - Neuropsychiatr. Genet.*, **114**, 584–591.
- 456 32. Sun, M., Latourelle, J. C., Wooten, G. F., et al. (2006) Influence of heterozygosity for parkin
457 mutation on onset age in familial Parkinson disease: the GenePD study. *Arch. Neurol.*, **63**,
458 826–832.
- 459 33. Nalls, M. a, Bras, J., Hernandez, D. G., et al. (2014) NeuroX, a fast and efficient genotyping
460 platform for investigation of neurodegenerative diseases. *Neurobiol. Aging*, 1–6.
- 461 34. Wang, K., Li, M., Hadley, D., et al. (2007) PennCNV: An integrated hidden Markov model
462 designed for high-resolution copy number variation detection in whole-genome SNP
463 genotyping data. *Genome Res.*, **17**, 1665–1674.
- 464 35. Petitti, D. B. (1994) Meta-analysis, Decision Analysis and Cost-Effectiveness Analysis. *Meta-*
465 *analysis, Decision Analysis and Cost-Effectiveness Analysis*; Oxford University Press, New York,
466 (1994) .
- 467 36. Higgins, J. P. T. and Thompson, S. G. (2002) Quantifying heterogeneity in a meta-analysis.
468 *Stat. Med.*, **21**, 1539–58.

469

470 LEGENDS TO TABLES

471 Table 1: Parkinson’s risk profiles associated with single heterozygous *PARK2* mutations.

472 LEGENDS TO FIGURES

473 Figure 1: Forest plot of the odds ratio (OR) of the Parkinson’s risk associated with heterozygous
474 *PARK2* mutations in three independent International Parkinson’s disease Genomics Consortium
475 (IPDGC) cohorts.

476 Figure 2: Forest plot of the odds ratio (OR) of the Parkinson's risk associated with heterozygous
477 *PARK2* mutations.

478 Figure 3: Forest plot of the odds ratio (OR) of the Parkinson's risk associated with heterozygous
479 *PARK2* single nucleotide variant (SNV) and copy number variant (CNV) carriers.

TABLES

Table 1: NeuroX Parkinson's risk profiles associated with single heterozygous *PARK2* mutations.

Type	N (Freq)		OR (95% CI)	P _{log}	AAO	Coeff (95% CI)	P _{reg}
	Controls	PD					
All							
With	37 (0.6%)	66 (1.0%)	1.55 (1.03, 2.32)	0.035	56.4	-5.04 (-8.34, -1.75)	0.003
Without	5,656	6,486			61.4		
CNV							
With	4 (0.1%)	11 (0.2%)	2.53 (0.80, 7.99)	0.113	58.4	-3.11 (-11.15, 4.93)	0.449
Without	5,689	6,541			61.4		
SNV							
With	33 (0.6%)	55 (0.8%)	1.43 (0.92, 2.21)	0.108	56.0	-5.41 (-9.02, -1.81)	0.003
Without	5,660	6,497			61.4		

Key: AAO, age at onset; CI, confidence interval; CNV, copy number variant; Coeff, linear regression coefficient correcting for gender and principal components 1-4; Freq, frequency; N, number of samples; OR, odds ratio correcting for gender and principal components 1-4; *PARK2*, *Parkin* (NM_013988 and NM_004562); PD, Parkinson's cases; SNV, single nucleotide variant.

FIGURES

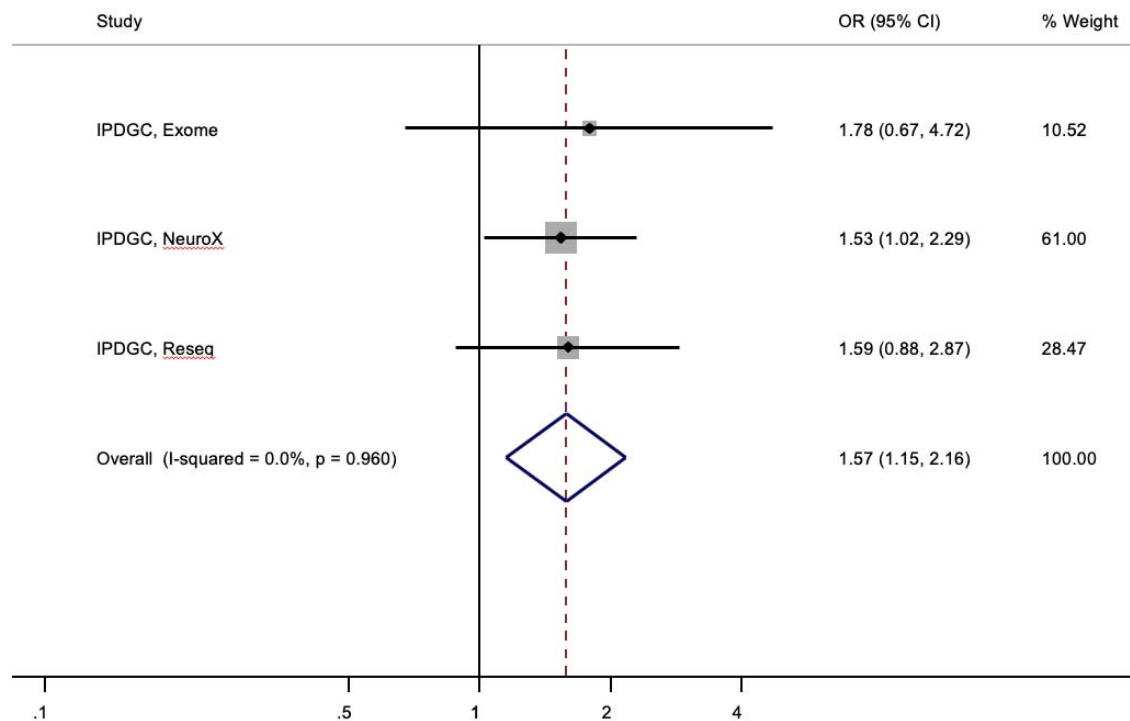


Figure 1: Forest plot of the odds ratio (OR) of the Parkinson's risk associated with heterozygous *PARK2* mutations in three independent International Parkinson's disease Genomics Consortium (IPDGC) cohorts.

Boxes denote OR point estimates, their areas proportional to the inverse variance weight of the estimate. Horizontal lines represent 95% CIs. Vertical dashed line represents pooled OR point estimates. Key: CI, confidence intervals; *PARK2*, *Parkin* (NM_013988 and NM_004562); %, percentage; Reseq, resequencing.

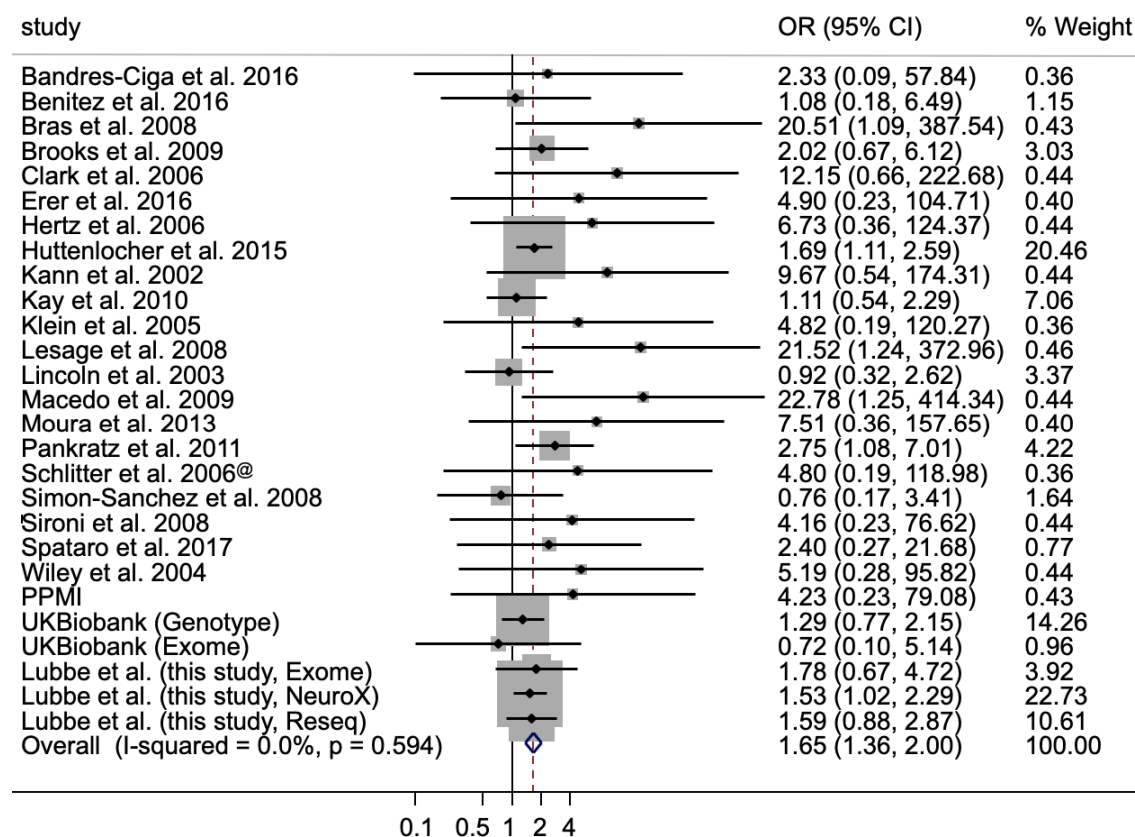


Figure 2: Forest plot of the odds ratio (OR) of the Parkinson's risk associated with heterozygous *PARK2* mutations.

Boxes denote OR point estimates, their areas proportional to the inverse variance weight of the estimate.

Horizontal lines represent 95% CIs. Vertical dashed line represents pooled OR point estimates. Key: CI,

confidence intervals; *PARK2*, *Parkin* (NM_013988 and NM_004562); %, percentage; * German samples only.

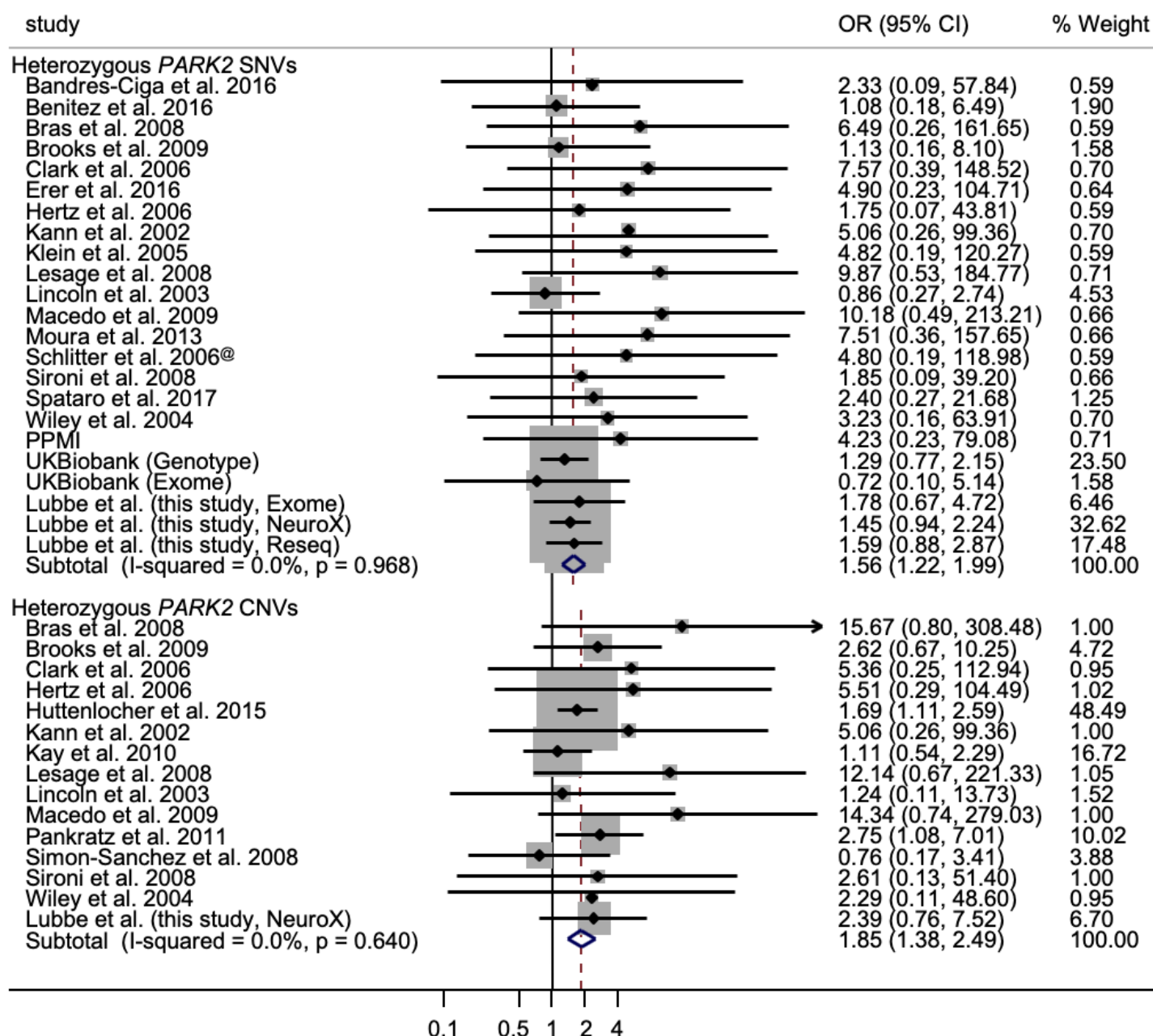


Figure 3: Forest plot of the odds ratio (OR) of the Parkinson's risk associated with heterozygous *PARK2* single nucleotide variant (SNV) and copy number variant (CNV) carriers.

Boxes denote OR point estimates, their areas proportional to the inverse variance weight of the estimate. Horizontal lines represent 95% CIs. Vertical dashed line represents pooled OR point estimates. Key: CI, confidence intervals; IPDGC, International Parkinson's disease Genomics Consortium; *PARK2*, *Parkin* (NM_013988 and NM_004562); %, percentage; PPMI, Parkinson's Progression Markers Initiative; Reseq, Resequencing; * German samples only.

ABBREVIATIONS

%	Percent
AAO	Age at onset
AR	Autosomal recessive
c.	coding DNA reference sequence
C1-4	Principal component 1 to 4
CI	Confidence interval
CNV	Copy number variant
Coeff	β -Coefficient
dbGaP	Database of Genotypes and Phenotypes
DNA	Deoxyribonucleic acid
EGA	European Genome-phenome Archive
EOPD	Early-onset Parkinson's
<i>FBXO7</i>	F-box protein 7
Freq	Frequency
I^2	Proportion of total variation caused by heterogeneity
IPDGC	International Parkinson's Disease Genomics Consortium
<i>LRRK2</i>	Leucine-rich repeat kinase 2
MLPA	Multiplex ligation-dependent probe amplification
MRC-Holland	Microbiology Research Centre Holland
N	Number
NM_*	Messenger RNA sequence identifier
OMIM	Online Mendelian Inheritance in Man
OR	Odds ratio
P	P-value
p.	protein reference sequence
<i>PARK2</i>	Parkin
<i>PARK7</i>	Parkinsonism associated deglycase or <i>DJ-1</i>
PD	Parkinson's
PD-Monoallelic	Parkinson's cases harbouring a single heterozygous <i>PARK2</i> mutation
P_{het}	Cochran's Q-statistic test for heterogeneity P-value
<i>PINK1</i>	PTEN-Induced putative kinase 1

$P_{\log}$	Logistic regression P-value
PPMI	Parkinson's Progression Markers Initiative
P_{reg}	Linear regression P-value
Reseq	Resequencing
<i>SNCA</i>	α -Synuclein
SNV	Single nucleotide variant
<i>VPS35</i>	Vacuolar protein sorting 35, yeast, homolog of
vs.	Versus