Insights on Genetic and Environmental Factors in Parkinson's Disease from a regional Swedish Case-Control Cohort

Kajsa Brolin[1*], Sara Bandres-Ciga[2], Cornelis Blauwendraat[2], Håkan Widner[3,4], Per Odin[3,4], Oskar Hansson[5,6], Andreas Puschmann[3,4], Maria Swanberg[1]

[1]Lund University, Translational Neurogenetics Unit, Department of Experimental Medical Science, Lund, Sweden, [2]Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA, [3]Lund University, Department of Clinical Sciences Lund, Neurology, Sweden, [4]Department of Neurology, Skåne University Hospital, Sweden, [5]Clinical Memory Research Unit, Department of Clinical Sciences Malmö, Lund University, Sweden, [6]Memory Clinic, Skåne University Hospital, Malmö, Sweden

*Corresponding author: Kajsa Brolin, kajsa.brolin@med.lu.se, Translational Neurogenetics, BMC A10, 221 84, Lund, Sweden

Abstract

BACKGROUND: Risk factors for Parkinson's disease (PD) can be more or less relevant to a population due to population-specific genetic architecture, local lifestyle habits, and environmental exposures. Therefore, it is essential to study PD at a local, regional, and continental scale in order to increase the knowledge on disease etiology.

OBJECTIVE: We aimed to investigate the contribution of genetic and environmental factors to PD in a new Swedish case-control cohort.

METHODS: PD patients (n=929) and matched population-based controls (n=935) from the southernmost county in Sweden were included in the cohort. Information on environmental exposures was obtained using questionnaires at inclusion. Genetic analyses included a genome-wide association study (GWAS), haplotype assessment, and a risk profile analysis using cumulative genetic risk scores.

RESULTS: The cohort is a representative PD case-control cohort (64% men, mean age at diagnosis=67 years, median Hoehn and Yahr score=2.0), in which previously reported associations between PD and environmental factors, such as tobacco, could be confirmed. We describe the first GWAS of PD solely composed of PD patients from Sweden, and confirm associations to well-established risk alleles in *SNCA*. In addition, we nominate an unconfirmed and potentially population-specific genome-wide significant association in the *PLPP4* locus (rs12771445).

CONCLUSIONS: This work provides an in-depth description of a new PD case-control cohort from southern Sweden, giving insights into environmental and genetic risk factors of PD in the Swedish population.

MESH keywords:

Parkinson Disease, Sweden, Case-Control Studies, Risk Factors, Genome-Wide Association Study, Haplotypes

48 # Introduction

49

50 Parkinson's disease (PD) has been reported to be the fastest growing neurological disorder
51 worldwide in terms of prevalence, disability, and deaths [1, 2]. Currently available treatments
52 only address symptoms, and the burden of PD will increase substantially if no disease-
53 modifying therapeutic is developed.

54

55 PD can be categorized into monogenic and idiopathic (multifactorial) based on the contribution
56 of genetics to disease risk. In monogenic PD, rare DNA variants with large effect sizes but
57 varying penetrance cause the disease [3]. Although such variants in *SNCA*, *LRRK2*, *VPS35*,
58 *PRKN*, *PINK1, GBA,* and *DJ-1* [4-9] contribute to a small proportion of cases, these genes have
59 provided valuable insight into the molecular mechanisms underlying PD etiology and
60 pathogenesis, as well as an overlapping etiology between monogenic and idiopathic PD.
61 Missense mutations and copy number variants (CNVs) of *SNCA*, encoding a-synuclein, have
62 been identified to cause monogenic PD, whereas common *SNCA* variants have been associated
63 with an increased risk of idiopathic PD [5]. In idiopathic PD, accounting for more than 95% of
64 all PD cases, variants with relatively small effect sizes in combination with environmental
65 exposures influence an individual's disease risk [10, 11]. The latest large-scale genome-wide
66 association study (GWAS) meta-analysis identified 90 risk loci associated with PD in
67 individuals of European ancestry [12]. These variants were estimated to explain 16-36% of the
68 heritable PD risk, indicating that many genetic risk factors for PD remain to be discovered [12].
69 Additionally, two new risk loci have been reported to be associated with PD risk in the Asian
70 population [13].

71

Larger sample sizes in GWAS could pave the way for the discovery of additional risk variants and increase the proportion of explained heritability [14, 15]. However, it has also been suggested that the "missing heritability" could be explained by the presence of gene-gene and gene-environment interactions [14]. Commonly reported environmental risk factors for PD include smoking and exposure to pesticides, two exposures that can vary a lot between population. Further understanding environmental risk factors for PD in different populations are hence of great importance to further elucidate the pathogenesis of the disease [10, 16, 17].

There is a strong population bias towards individuals of European ancestry in genetic research of human diseases [18, 19], but initiatives are taken to study PD genetic risk in other populations [20, 21]. In addition to the need of a widened scope of PD genetic research on a global level, narrowed studies are also of interest, since heterogeneity across cohorts from different countries or regions may mask genetic associations specific to sub-population [22].

The Swedish population has not yet been represented in PD GWAS meta-analyses or population-specific GWAS aside from a PD GWAS performed on a mixed Scandinavian population from Norway and Sweden [23]. The Swedish population has been reported to be genetically similar to the nearby population of Norway [24, 25]. However, it has also been reported that Swedish genomes contain a substantial amount of genetic variation and genetic variants that are not represented in other European populations [26], and that there is a large genetic difference even within the Swedish population, particularly between the south and north of Sweden [24, 26]. This highlights the importance of analyzing different European populations, and emphasize the importance of regional matching of cases and controls in GWAS [24, 27].

97 Here, we provide a comprehensive description of a novel, well-defined, and matched PD case-
98 control cohort from southern Sweden in which we describe, to our knowledge, the first GWAS
99 of PD composed solely of individuals from Sweden. This study provides insights into
100 environmental and genetic risk factors of PD in the Swedish population and will contribute to
101 the understanding of population-specific environmental and genetic risk in PD.
102

103 Materials and Methods

104

105 Ethical consent

106 Ethical permission for the MultiPark's biobank sample collection (MPBC) was approved by
107 the regional ethics review board of Lund (2013/509). All participants gave informed written
108 consent at study enrollment.
109

110 Patient and population-based control inclusion

111 Individuals primarily diagnosed with PD (*International Classification of Disease*, *tenth*
112 *revision (ICD-10-SE)* code G20.9) in the southernmost province of Sweden, Scania, were
113 assessed for eligibility (Figure S1). Inclusion criteria were a PD diagnosis and the ability to
114 visit one of the clinics for sampling. In total, 2,119 PD patients were invited to participate
115 between November 2014 and July 2017, whereof 1,011 were included (inclusion rate 48.3%).
116 For each patient, population-based controls matched by date of birth, sex, and residential area
117 were randomly selected from the Swedish Population Register and invited between March
118 2015 and April 2018. A total of 1,001 individuals consented and completed data collection
119 (inclusion rate 18.5%). Exclusion criteria were a PD diagnosis and the inability to visit a center
120 for sampling. Out of the completed controls, 953 were unique matchings, resulting in a matched
121 control for 953 PD patients (94.3%). All analyses were performed in a set of 929 patients and

122 935 controls passing all quality control (QC) steps unless stated otherwise. A flowchart
123 detailing the study participant inclusion process can be found in Figure S2.
124
125 Sample and data collection, storage, and management
126 Blood samples were handled by the Clinical Chemistry Unit at the University Hospital in
127 Scania (SUS) and stored at -80°C within a central biobanking facility (Biobank Syd: BD47).
128 Whole blood was collected in EDTA-tubes (Becton Dickinson (BD) Vacutainer®, New Jersey,
129 US) for DNA extraction. Additional samples were collected in PAXgene blood RNA tubes
130 (QIAGEN®, Hilden, Germany), EDTA-tubes (BD Vacutainer®), sodium-heparin tubes (BD
131 Vacutainer®, New Jersey, US) and a serum-separating tube (BD Vacutainer® SST™, New
132 Jersey, US) to enable future studies of RNA and biomarkers in PD. When difficulties emerged
133 during blood sample collection, EDTA-tubes for DNA extraction were prioritized. Serum and
134 plasma samples are stored at -80°C in a 96-well format of 200 µl aliquots.
135
136 All participants filled in questionnaires covering basic demographic data, lifestyle habits,
137 family history of PD, comorbid diseases, environmental exposures, medications, health status,
138 and perceived motor- and non-motor symptoms. Data from questionnaires was manually
139 entered into the web-based application REDCap [28]. Additional data for the PD rating tools
140 Hoehn and Yahr (H&Y), Clinical Impression of Severity Index - Parkinson's Disease (CISI-
141 PD), and Parkinson's Disease Questionnaire - 8 (PDQ-8), as well as information on age at
142 diagnosis (AAD) was retrieved from the Swedish Parkinson's registry
143 (https://www.neuroreg.se/parkinsons-sjukdom/).
144
145
146

147 Statistical analyses of epidemiological data

148 Epidemiologic data was analyzed using R version 4.0.0 [29]. Demographic characteristics were
149 summarized as frequency and percentage for categorical variables. For continuous variables,
150 mean and standard deviation (SD) were used for variables displaying normal distribution
151 whereas median and interquartile range was reported for non-normally distributed variables.
152 The range for continuous variables was used to demonstrate the heterogeneity in the cohort.
153 Body mass index (BMI) was calculated as weight (kg)/height $(m)^2$. For quality-of-life
154 assessment, EuroQol five-dimension-3-level (EQ-5D-3L) was used, covering mobility, self-
155 care, usual activities, pain/discomfort, and anxiety/depression. Time trade-off (TTO) and the
156 visual analogue scale (VAS) index were calculated using scale value sets for Sweden [30]. In
157 TTO valuation, individuals were asked to indicate how many years in full health that would be
158 of equal value to 10 years in their current health state (divided by 10). Full health hence
159 corresponds to 10 years/10=1. In VAS valuation, respondents rated their overall health between
160 0 (worst imaginable health) and 100 (best imaginable health) [31]. Additionally. the occurrence
161 of PD symptoms at study inclusion were compared between the patients and controls using
162 Pearson's chi square test with a Bonferroni corrected threshold for statistical significance of
163 p=0.001.

164

165 To test for associations between potential risk factors and PD status, both unadjusted and
166 multivariate logistic regression analyses were used. Risk factors were selected based on meta-
167 analysis literature reviews [16, 17]. Due to the cross-sectional case-control study design, we
168 selected variables from the questionnaire related to past exposures, excluding variables related
169 to exposures at time or near time of inclusion. Directed acyclic graphs (DAGs) were made for
170 all exposures of interest using the R package 'dagitty' to help identify confounding variables
171 to adjust for in the multivariate analyses (Figure S3) [32]. Confounding pathways with the least

172 amount of missing data were prioritized and complete-case analyses were run. Odds ratios
173 (OR), and 95% confidence intervals (CI) were estimated and plotted in a forest plot.
174
175 We further evaluated variables for BMI and comorbidities at inclusion, along with alcohol and
176 red wine consumption and physical exercise the year prior to inclusion, and ibuprofen use the
177 past two weeks prior to inclusion, using logistic regression analyses, adjusting for age and sex.
178 We acknowledge the risk of confounding factors but the interest was solely to perform
179 observational analyses. Complete-case analyses were run and ORs and 95% CI were estimated
180 and plotted in a forest plot. A total of 35 comorbidities were not evaluated for association due
181 to either too few study participants reporting a certain diagnosis or a lack of difference between
182 the occurrence of the diagnosis between the patient and control group.
183
184 DNA extraction and Genotyping
185 DNA extraction was done from whole blood at BD47 and at LGC Biosearch Technologies
186 (UK, GEN-9300-120). LGC Biosearch Technologies (Germany) also performed genotyping
187 and subsequent "Basic BioIT" with technical QC and genotype clustering. The no-call GenCall
188 threshold was set to the standard score cutoff for Infinium data of 0.15. Matched case-control
189 samples were genotyped using the Infinium Global Screening Array-24 v.2.0 with the Multi-
190 Disease drop in panel v.2.0 (GSAMD-24v2.0) containing 712,189 variants annotated to the
191 Genome Reference Consortium Human Build 37 (GRCh37).
192
193 Genotype quality control and Imputation
194 QC analysis was performed in PLINK 1.9 [33, 34]. Samples were excluded if the call rate was
195 <95%, if excess heterozygosity was detected (estimated by a F statistic >0.15 or <-0.15), or if
196 the genetic sex (determined from X-chromosome heterogeneity) differed from the reported sex.

197 Principal component analysis (PCA) was used to identify and exclude ancestry outliers. For
198 this, the genotype data was merged with the HapMap phase 3 data set [35], and samples were
199 determined as having European ancestry if they clustered 6 SD (+/-) around the combined
200 population mean of the populations: Utah residents with Northern and Western European
201 ancestry (CEU) and Toscani in Italia (TSI) (Figure S4). Samples from closely related
202 individuals, defined as sharing more than 12.5% of alleles, were also excluded. Following
203 sample exclusion, 929 patients and 935 controls remained. Variant level QC was subsequently
204 done, excluding SNPs with a missingness rate >5%, Hardy-Weinberg equilibrium p-value <1E-
205 4, or minor allele frequency (MAF) <1%. SNPs with differences in genotyping rate for patients
206 and controls, and genotypes missing by haplotype, were also excluded for p-value <1E-4. SNPs
207 passing QC were 505,005 with a total genotyping rate of 99.9%. Genomic imputation was
208 performed on the Michigan Imputation Server [36] using the Minimac4 imputation software,
209 Eagle v2.4 for phasing [37], and the Haplotype Reference Consortium (HRC) r1.1 2016
210 European population as the reference population (http://www.haplotype-reference-
211 consortium.org/) [38]. Post-imputation QC was performed, and variants with MAF<5% and/or
212 poor imputation quality, set as Rsq <0.3, were excluded. The number of imputed variants after
213 filtering was 5,445,841.
214
215 Power calculation for GWAS analyses
216 GWAS power calculation was performed using the online Genetic Association Study (GAS)
217 Power Calculator (http://csg.sph.umich.edu/abecasis/gas_power_calculator/). In order to reach
218 80% power at a GWAS-significance level (alpha=5E-08) with a disease prevalence in the
219 general Swedish population of 0.2%, an allele frequency of >20% and an OR>1.7 would be
220 needed.
221

222 Genome wide association study (GWAS) vs PD risk and age at diagnosis (AAD)

223 A GWAS for PD risk was performed in a logistic regression model adjusted for sex, age (at
224 study inclusion), and the first 5 principal components (PCs). The number of PCs used in the
225 analyses was determined by a scree plot (Figure S5). A linear regression model was used to
226 investigate the impact of genetic variation on the AAD, adjusted for sex and PC1-5. Data on
227 AAD was available for 792 of the 929 PD patients (85.3%). Summary statistics was generated
228 using RVTESTS (version 2.1.0) [39]. Quantile-quantile (QQ)-plots and Manhattan plots were
229 generated in R, version 4.0.0 [29]. Subsequently, regions of interest were visualized using
230 summary statistics data in the LocalZoom tool (https://my.locuszoom.org) with a flanking size
231 of 100 kb from the gene of interest.

232

233 Haplotype analysis of *PLPP4*

234 Genotyped variants in the *PLPP4* region with a window of ±100 kb (*PLPP4* coordinates
235 on GRCh37; chr10:122216466-122349376, UCSC) was extracted and annotated using
236 ANNOVAR [40]. Haplotype blocks in the region were estimated in PLINK v1.9 using the
237 default parameters wherein pairwise LD was calculated for variants within 500 kb [33, 34].
238 PLINK v1.9 estimates haplotype blocks following the default procedure in Haploview [41, 42].
239 Association between haplotype blocks in *PLPP4* and PD status was analyzed with haplo.stats
240 in R. The regression analyses were adjusted for age, sex, and PC1-5, and additionally for
241 tobacco use. The adjustment for tobacco use was done because of a report of *PLPP4* being
242 associated with smoking cessation [43], and multiple reports of smoking being associated with
243 PD. To assimilate the effect of other nicotine products, such as snus (moist tobacco, commonly
244 consumed in Sweden), the variable "Tobacco" was used. Only haplotypes with a carrier
245 frequency of >1% were included in the analysis. A Bonferroni corrected threshold for statistical

246 significance of p=0.05/16=0.003. was set. The results were visualized using LocusZoom
247 (https://my.locuszoom.org) and Haploview v4.1 [44].
248
249 Analysis of the joint effect of variants in *PLPP4* on PD risk
250 The contribution of rare variants on PD risk was examined by the joint effect of multiple
251 variants in the *PLPP4* region. We separately analyzed genotyped and imputed variants with an
252 imputation quality score of Rsq >0.8 for imputed variants. The sequence kernel association test
253 (SKAT) was used since only 2.6% of the variants in *PLPP4* are coding variants and the SKAT
254 is powerful when a large fraction of variants in a region are noncoding. The test aggregates
255 associations between variants and phenotype and allows for variant interactions [45]. The
256 analyses were performed in RVTESTS (version 2.1.0) [39] using default parameters and
257 adjusting for age, sex and PC1-5. The test was performed both for all available variants as well
258 as in only coding variants with two different maximum MAF thresholds of <1% and <5%. To
259 adjust for multiple comparisons, Bonferroni corrected p-value thresholds was applied for all
260 analyses.
261
262 Risk profile analysis
263 A risk profile analysis for PD and AAD was performed using the imputed data passing QC. A
264 cumulative genetic risk score (GRS) was calculated in PLINK v1.9 using publicly available
265 effect estimates (beta coefficients) of 90 SNPs associated with PD in the largest published
266 meta-analysis of PD GWAS to date [12]. For each SNPs, the allele dosage was multiplied by
267 the beta coefficient, giving a greater weight to alleles with higher risk estimates. Risk allele
268 dosage was then summed across all variants to generate a GRS for each study participant.
269 Subsequently, the GRSs were standardized to Z-scores using the control group as the reference
270 group for PD status and the patient group for the AAD analysis. Following standardization, a

271 Z score of 1 can be interpreted to be equivalent to one SD increase in the GRS from the
272 reference group GRS mean. The study participants were divided into quartiles based on their
273 GRS, and logistic regression analysis was performed to investigate associations and estimate
274 the OR between PD status and quartile group. Covariates were selected from a previously
275 published PD classification [46] and included PD family history (yes/no), age at inclusion, sex,
276 and additionally PC1-5. For the risk profile analysis for AAD, a linear regression model was
277 used, adjusted for sex, PD family history, and PC1-5. Plots were generated in R, version 4.0.0
278 [29].
279
280 Code for all analyses will be available at https://github.com/KajBro/MPBC, GWAS summary
281 statistics will be available online.
282

283 Results

284

285 Characteristics of the MPBC cohort

286 Demographics and characteristics of the MPBC cohort are summarized in Table 1. An expected
287 male predominance was observed with a sex ratio of 2:1. Case-control matching was successful
288 for year of birth and sex, with average birth year 1944 and 36% women in both groups. A one-
289 year difference for the average age at inclusion was observed between patients (71 years) and
290 controls (72 years) due to control matching occurring after patient inclusion. Swedish ancestry
291 (defined as self-reporting that both parents were born in Sweden) was similar in patients and
292 controls (87.9% vs 86.1%). Similar distributions for highest completed education and marital
293 status were also seen. According to self-reported information on health status using the EQ-
294 5D-3L instrument [30], patients had an index of 0.8 on the TTO scale compared to 0.9 for the
295 controls. The patients also rated their own overall health lower than the controls on the VAS

296 index scale (index 68.9±16.8 vs 82.0±9.7, Table 1). As a reference, the general Swedish
297 population has an estimated mean TTO index of 0.9 and a VAS index of 79.5 [30]. As expected,
298 all PD symptoms were more common in the patient group than the control group (Table S3).
299 The most common self-reported motor-symptoms were muscle stiffness (72.7%) followed by
300 slowness of movements (72.4%). The most commonly reported non-motor symptom for both
301 patients and controls was nocturia (71.6% vs 58.4%).
302
303 We further evaluated the health status and AAD among male and female PD patients (Table
304 2). Median AAD was identical for both sexes (67.0 years), although with a wide range (men
305 29.0-84.0 years, women 35.0-89.0 years). The median time since diagnosis at inclusion was
306 4.0 years for both sexes and varied from diagnosis the same year to 36.0 years for men and
307 33.0 years for women. Health status at the time of inclusion was evaluated using the H&Y
308 scale, CISI-PD, and PDQ-8. Both sexes had a median of 2.0 on the H&Y, corresponding to
309 bilateral involvement without impairment of balance. Similar total score values were reported
310 by men and women on the CISI-PD scale (median 5.0, range 0-24). Women had a slightly
311 higher median PDQ-8 total score compared to men (7.0 vs 6.0 out of 32).
312
313 Epidemiological characteristics of PD in Sweden
314 We further set out to evaluate previously confirmed risk factors for PD in a regional, Swedish
315 context (Figure 1, Table S1). A positive family history (at least one relative diagnosed with
316 PD) was overrepresented among PD patients compared to controls (20% vs 11%).
317 Consequently, having a relative diagnosed with PD doubled the risk of having PD (OR=2.00,
318 95% CI 1.51-2.67). An association between pesticide exposure and PD was confirmed at an
319 OR of 2.26 (95% CI 1.39-3.72). A significant association was also seen between PD and a
320 history of head trauma (OR=1.30, 95% CI 1.08-1.58). However, no association between loss

321 of consciousness and PD was observed among study participants who reported a history of
322 head trauma. A slightly higher OR was seen for a higher BMI at the age of 20 years (1.05, 95%
323 CI 1.01-1.09), indicating that a one unit increase in BMI increased the risk of having PD by
324 5%. However, no difference was observed for the highest reported BMI. We observed an
325 inverse association between having ever smoked and PD, but the association was not
326 statistically significant after adjustment for confounders (ever- vs never-smoking OR=0.82,
327 95% CI 0.67-1.01). A statistically significant inverse association between snus and PD was
328 observed also after adjusting for confounders (OR=0.53, 95% CI 0.38-0.73). An inverse
329 association was also observed for the combined effect of tobacco products on PD (adjusted
330 OR=0.72, 95% CI 0.59-0.88). Moreover, a trend of lower OR for PD with increasing amount
331 of coffee drinking at all investigated age groups was observed, where drinking >5 cups of
332 coffee per day was inversely associated with PD (OR 0.36-0.52, 95% CI 0.18-0.86).
333
334 Additionally, variables related to exposure/use within the past year were examined to describe
335 potential differences between the age- and sex-matched groups at inclusion (Figure 2, Table
336 S2; Table S3). The patients consumed less alcohol and engaged in less physical activity both
337 in regard to level of physical activity (regular activity level vs nothing OR=0.26, 95% CI 0.18-
338 0.38) and hours per week (≥5 hours/week vs nothing, OR=0.21, 95% CI 0.13-0.33). No
339 significant association between PD and BMI at inclusion was observed (OR=0.99, 95% CI
340 0.96-1.00). For comorbidities, statistically significant inverse associations between PD and a
341 diagnosis of hyperlipidemia (OR=0.51, 95% CI 0.40-0.64), hypertension (OR=0.57, 95% CI
342 0.47-0.69), and osteoarthritis (OR=0.66, 95% CI 0.53-0.82) were seen. A significant
343 association with PD was observed for a diagnosis of depression (OR=1.89, 95% CI 1.39-2.59),
344 back pain (OR=1.56, 95% CI 1.15-2.13), and bowel problems, where bowel problems (defined
345 as having constipation or diarrhea that required treatment on a regular basis) almost quadrupled

346 the likelihood of having PD (OR=3.93, 95% CI 2.66-5.97). We observed an OR of 0.53 (95%
347 CI 0.41-0.68) for the use of ibuprofen <2 times/week compared to never. However, the
348 association was lost for ibuprofen use ≥2 times/week. No associations were observed for the
349 use of other nonsteroidal anti-inflammatory drugs (NSAIDs), such as diclofenac or naproxen
350 (data not shown).
351
352 GWAS of PD risk nominated a novel genome-wide significant variant in the *PLPP4* locus
353 The variant rs356182 in *SNCA* was reported in the latest GWAS meta-analysis as one of the
354 90 risk loci associated with PD in a cohort of European ancestry [12]. In our GWA analysis for
355 PD risk, this and another three variants in the *SNCA* locus were observed at low p-values
356 (rs356182, beta=-0.38, SE=0.07, p=5.64E-08; rs356203, beta=-0.36, SE=0.07, p=1.27E-07;
357 rs356220, beta=0.36, SE=0.07, p=1.39E-07; rs356219, beta=-0.36, SE=0.07, p=2.32E-07), but
358 none reached genome-wide significance (Table S4). In addition to the *SNCA* variants, we found
359 associations for 23 of the 90 previously reported PD risk loci in this Swedish cohort at an
360 uncorrected p-value of 0.05 (Table S5). Results from PCA showing the population structure as
361 compared to the HapMap3 reference populations, revealed that this Swedish cohort cluster near
362 the other European population but the centers of the clusters differ slightly, indicating
363 differences in population structures that could result in population-specific risk alleles (Figure
364 S6). The GWA analysis also revealed a novel genome-wide significant association signal in
365 the *PLPP4* locus, rs12771445 (10:122318147, beta=-0.44, SE=0.07, p=1.30E-08), as being
366 associated with PD in this Swedish cohort (Figure 3, Table S4). This association has, to our
367 knowledge, not been reported in any other cohort (Table S4) [12]. The significant variant was
368 an intron variant in the gene Phospholipid Phosphatase 4 (*PLPP4*) and was imputed with
369 Rsq=0.99 and MAF=0.31 (Table S4). The MAF is similar to the value observed in the Swedish
370 reference population, SweGen, of 0.32 [26]. The estimated MAF was 0.27 in the patient group

371 and 0.36 in the control group, indicating a potential protective effect of the minor allele (T)
372 with an OR of 0.64. To further display the association signals relative to genomic position and
373 local LD, locus zoom plots for *PLPP4* rs12771445 and *SNCA* rs356182 were produced (Figure
374 4 and Figure S10). Variants in *SNCA* have been reported to also be associated with PD AAO
375 [47]. We performed a GWA for the related variable AAD of PD in the cohort. However, we
376 did not detect any genome-wide statistically significant associations and could not replicate
377 previous reported variants associated with PD AAO at low p-values, likely due to smaller
378 sample size (792 vs 28,568) [47], and limited statistical power (Figure S11 and S12).
379
380 In order to analyze more genetic variants, imputation was done using the beta-version of the
381 TOPMed Imputation Reference panel [48]. Using the same post-imputation QC, 6,214,098
382 variants were generated and analyzed in a GWAS for PD risk (Figure S8 and S9). Two variants
383 passed genome-wide significance, rs12772937 (chr10:122318146, beta=-0.40, SE=0.07,
384 p=2.48E-08) and our previously identified rs12771445 (10:122318147, beta=-0.40, SE=0.07,
385 p=2.48E-08). Since the TOPMed panel was still in the beta stage as of our analysis, we
386 continued using the data imputed with the HRC version r1.1 2016 panel.
387
388 Haplotype analysis of *PLPP4*
389 Because of the identification of the novel genome-wide signal (rs12771445) associated with
390 PD in this cohort, we further looked at haplotypes among the genotyped variants in the nearest
391 gene, *PLPP4* (± 100 kb), to capture the genetic variance in the region. We identified 92
392 genotyped variants passing QC in the region, where 32 were in *PLPP4*, all intron variants. The
393 median GenCall scores of these variants was 0.87 (IQR: 0.83–0.91) (Figure S13). A total of 45
394 different haplotypes in 12 different blocks (H1-H12) in the region was identified (Table 3,
395 Figure S14). Statistically significant association with PD was observed for a haplotype in one

396 block (H6) (p=2.45E-05), both before and after adjustment. The haplotype block spanned over
397 41.9 kB (chr10:122302038-122343950) and consisted of five genetic variants and six different
398 haplotypes. The rs978854(G)-rs7910507(G)-rs10886711(G)-rs11199417(G)-rs10886717(A)
399 haplotype was statistically significantly inversely associated with PD at an OR of 0.69 (95%
400 CI 0.59-0.82, p=2.54E-05) and with a MAF=0.24 for patients and 0.30 for controls.
401
402 Analysis of the joint effect of variants in *PLPP4* on PD risk
403 We further evaluated the joint effect of more rare variants (MAF <5%) in *PLPP4* on PD risk
404 using both genotyped and imputed data. No coding variants in the gene were identified in the
405 genotyped data and only one was identified in the imputed data (Table 5). When focusing on
406 all variants, an association in the genotyped data at MAF<5% was observed (p=0.002), passing
407 Bonferroni correction. A trend of a potential joint effect of variants with a MAF<5% was still
408 observed in the imputed data (p=0.005). However, it did not reach multiple testing corrections
409 (p-value threshold=0.05/628=7.96E-05). Only one variant with MAF<1% was present in the
410 genotyped data and no statistically significant association was observed for variants with
411 MAF<1% in either the genotyped or the imputed data.
412
413 Genetic PD risk profile
414 The PD risk profile analysis showed that the Z-standardized GRS from Nalls et al was
415 associated with PD status also in this Swedish case-control cohort [12]. For each SD increase
416 from the reference mean value, the OR increased by 1.8 (beta=0.59, SE=0.05, p<2.22E-16)
417 (Figure 5). Study participants in the highest GRS quartile were estimated to be 4.7 times more
418 likely to have PD compared to the participants in the lowest quartile using an adjusted model.
419 There was also an inverse association between the GRS and AAD, where one SD increase in

420 the Z-standardized GRS was associated with approximately one-year earlier AAD (beta=-0.97,
421 SE=0.36, p=0.007, adjusted r^2=0.011).
422

423 Discussion

424

425 This study is, to our knowledge, the largest case-control study of PD performed in a Swedish
426 cohort. The study includes a well-defined recruitment of matched study participants from a
427 specific region in southern Sweden. Multiple previously reported environmental and genetic
428 risk factors were confirmed to affect PD risk. Interestingly, a novel genome-wide significant
429 association with PD risk at the *PLPP4* locus was observed, both in the GWA and the haplotype
430 analysis but further validation of the nominated variant is needed.
431

432 The MPBC cohort demonstrates similar characteristics compared to other PD cohorts. At
433 inclusion, the patients had a relatively short disease duration (average 4 years), which was
434 reflected in the PD rating scales with an average score of 2 on the H&Y scale and low scores
435 on the CISI-PD and PDQ-8 scales. This can be explained by the study inclusion design where
436 patients with advanced PD, who could not visit a neurology clinic, were not eligible. For self-
437 reported PD symptoms, hypokinesia/bradykinesia was most common, occurring in 72% of the
438 patients. Non-motor symptoms occurring at later stages of the disease, such as delusions and
439 hallucinations were less frequent (6.0% and 16.8%), also highlighting the lower number of
440 patients with advanced PD in the cohort [49]. The lack of an association between BMI at
441 inclusion and PD further highlight this. Weight loss has frequently been observed in PD
442 patients as a consequence of reduced energy intake in combination with an increased energy
443 expenditure due e.g. levodopa-induced dyskinesia [50]. It has been estimated that PD patients
444 lose 3.5 kg every eight years post-diagnosis and that levodopa-induced dyskinesia commonly

445 develops after 3-5 years [50, 51]. Approximately 92% of the patients in the cohort reported
446 using levodopa therapy.
447
448 Several previously reported environmental PD risk factors could be confirmed in this Swedish
449 cohort. Among those, an association between pesticide exposure and PD was observe, where
450 exposure to pesticides more than doubled the risk of having PD. The number of participants
451 exposed to pesticides was overrepresented among those who had reported farming and well
452 water consumption. However, no statistically significant association was observed between
453 farming or well water and PD in the cohort. Several pesticides have been reported to be linked
454 to PD, including paraquat, glyphosate, and pyrethroids [52]. Paraquat was banned in Sweden
455 in 1983 while glyphosates and pyrethroids are still used [53]. However, further actions need to
456 be taken to investigate which chemicals stand behind the observed association.
457
458 Having a relative with a PD diagnosis doubled the risk of having PD in the cohort. Among the
459 20% of patients that reported a positive family history, 8.9% had a first-degree relative with
460 PD. Other studies report similar numbers with 15-25% of PD patients having a positive family
461 history and 10-25% having a first-degree relative with PD [54]. It has been suggested that the
462 risk of PD increases shortly after a traumatic brain injury (TBI) and that a history of
463 concussions results in a higher risk of PD [55, 56]. We also observed a statistically significant
464 association between a history of head trauma and PD, but whether the association is causative
465 or a consequence of prodromal PD remains to be determined. Among participants reporting
466 head trauma, no association was observed between PD and loss of consciousness.
467
468 The inverse association between smoking and PD is well-known and has been reported in
469 numerous studies [17]. This was initially replicated in the cohort, but no longer statistically

470 significant after adjustment for confounders, which could be due to the common co-occurrence
471 of smoking and the use of snus. Interestingly though, a strong inverse association between the
472 Swedish moist tobacco snus and PD was observed also after adjustment. The amount of
473 nicotine reaching the blood when using snus is equivalent of that of cigarette smoking, and the
474 observation reduces the number of candidate compounds underlying the inverse association
475 between smoking and PD [57, 58]. It has previously been reported that non-smoking, snus-
476 using Swedish men had 60% lower risk of PD compared to men who had never used snus [58].
477 Snus is more frequently used than smoking tobacco among men in Sweden, where 18% use
478 snus and 7% smoke on a daily basis. Among Swedish women, 5% use snus and 7% smoke on
479 a daily basis [59]. We investigated the total effect of smoking and snus on PD risk and found
480 a statistically significant inverse association to PD also for the combined variable "Tobacco".
481 Our findings support previous reports and suggest either that components in tobacco leaves
482 influence biological processes underlying PD, or that there is a reverse causation between
483 tobacco use and PD. Nicotine has been suggested to have neuroprotective properties but it also
484 stimulates dopamine release from nigrostriatal dopaminergic terminals. Hence, the lower use
485 of tobacco products could be a consequence of a reduced nicotine-evoked dopamine release
486 among prodromal PD patients [60]. Caffeine has also been reported to have neuroprotective
487 effects in PD animal models, and several epidemiological studies have reported a relationship
488 between increased coffee consumption and decreased risk of developing PD [10]. We could
489 also observe a trend of lower OR for PD with higher coffee consumption in all investigated
490 age groups in this cohort.
491
492 The comorbidities hypertension, hyperlipidemia, and osteoarthritis were inversely associated
493 with PD in the cohort. Autonomic dysfunction can occur in PD, resulting in abnormalities in
494 blood pressure, and orthostatic hypotension is common [61]. In concordance with our results,

495 meta-analysis of PD risk factors has identified an inverse association between hypertension
496 and PD risk [16]. However, this was only observed in case-control studies, indicating that the
497 association might be a consequence of the disease. For hyperlipidemia, there are conflicting
498 reports on the association with PD but a meta-analysis supports an inverse association [62]. It
499 is possible that the use of lipid-lowering drugs (statins) contributes to the observed conflicting
500 associations [63]. The study participants were asked whether they had been diagnosed with
501 hyperlipidemia and, hence, it was likely that they used pharmacological treatments. Statins are
502 the most commonly prescribed lipid-lowering drugs in Sweden, and we cannot rule out that the
503 observed inverse association between PD and hyperlipidemia is a consequence of statin use.
504 Moreover, the PD patients were less likely to have a diagnosis of osteoarthritis (21% patients
505 vs 29% controls). Contrary to our observation, arthritis (no distinction was made between
506 rheumatoid arthritis and osteoarthritis) has been reported to be the most prevalent comorbidity
507 in PD, with almost 47% reported having the condition [64], indicating that our observation
508 could be a consequence of underdiagnosis of osteoarthritis among the PD patients [65].
509 Furthermore, associations between PD and the comorbidities depression, back pain, and bowel
510 problems were observed. Use of ibuprofen has been reported to be inversely associated with
511 PD in a dose-dependent manner [66]. We observed a reduced PD risk for the use of the NSAID
512 ibuprofen <2 times/week but not for ≥2 times/week and a recently published large study do not
513 support any evidence of a decreased PD incidence among NSAID users [67].
514
515 Here, we describe to our knowledge the first GWAS of PD composed solely of PD patients
516 from Sweden, specifically the southernmost region of Sweden. Although the relatively small
517 sample size is a limitation in this study, issues with population stratification are expected to be
518 lower due to region-specific study recruitment and, potentially, more homogenous ancestry.
519 The observed near genome-wide significance associated variants at the *SNCA* locus indicate

520 that this cohort is a well-designed case-control cohort of PD, and one of the variants was
521 reported among the 90 risk loci in the largest GWAS meta-analysis to date [12]. Another 23 of
522 the 90 risk loci were replicated at a nominal uncorrected p-value <0.05. Noticeable, this study
523 had insufficient statistical power to detect variants with low MAF or small effect size, and the
524 majority of 90 risk loci had an OR of 0.8-1.2 [12]. Insufficient statistical power is likely also
525 the reason for the lack of an identification of any loci associated with AAD. Two genome-wide
526 significant association signals have previously been reported to be linked to PD AAO, one at
527 the *SNCA* locus and one at *TMEM175* [47].
528
529 Interestingly, the GWA analysis identified a novel variant, rs12771445, associated with PD at
530 a genome-wide significance level. This variant has not been reported in the PD literature and,
531 hence, may be specific for the Swedish population. Although, it could be that this common
532 variant is a tagging a population-specific rare variant as the Swedish population appear to
533 contain a substantial number of genetic variants that are not represented in other European
534 populations [26]. However, the significant variant was imputed and we remain cautious in
535 drawing any conclusions. As expected, we further identified one haplotype in *PLPP4*
536 significantly associated with PD. The haplotype was more common among controls, indicating
537 a decreased risk of PD among carriers. Furthermore, we investigated the joint effect of rare
538 variants (MAF<5%) in *PLPP4* on PD. Due to the low coverage of coding variants, we adapted
539 the non-burden SKAT and observed an association with PD risk for the genotyped variants but
540 not for the imputed variants. Replication studies in larger Swedish cohorts are needed in order
541 to validate our findings of a potential association between *PLPP4* and PD in the Swedish
542 population.
543

544 The gene *PLPP4,* encodes for a phospholipid phosphatase that catalyzes the dephosphorylation
545 of bioactive lipid mediators such as phosphatidic acid (PA) [68]. Another regulator of PA
546 content, the diacylglycerol kinase, *DGKQ,* has previously been reported as a potential PD risk
547 factor [69, 70]. The role of lipids in PD is receiving greater attention and the question whether
548 if PD is a lipidopathy rather than a proteinopathy has been lifted [71]. Several reasons stand
549 behind this hypothesis, including the proposal that a-synuclein is a lipid-binding protein that
550 physiologically interacts with phospholipids and fatty acids (FAs) [71].
551
552 Although only 23 out of 90 previously reported risk variants were replicated in this Swedish
553 cohort, the cumulative GRS based on the 90 variants showed an almost five-fold higher risk
554 for PD in the highest quartile compared to the lowest quartile. Combining the GRS with
555 information on age, sex, family history of PD, and the University of Pennsylvania Smell
556 Identification Test (UPSIT) score (olfactory function) has been reported to further increase the
557 sensitivity and specificity of the PD status model [46]. Having access to all factors but the
558 UPSIT score, we added information on age, sex, family history of PD and PC1-5 to our model.
559 GRS alone gave an area under the roc curve (AUC) of 0.67 while the adjusted model gave an
560 AUC of 0.68. The non-observed difference for the two models could be due to the large
561 proportion of the variance explained by the UPSIT score in the reported prediction model
562 (63.1%) compared to the substantially lower values for GRS (13.6%), family history (11.4%),
563 gender (6.0%) and age (5.9%) [46]. The lack of association to PD in the Swedish cohort for
564 67/90 risk variants could also have an impact on the fit of the model applied in our cohort.
565 Although the model is functional enough to compare risk for disease status at a population
566 level, additional information would be needed for disease prediction. Interestingly, we
567 observed an association between the GRS and AAD despite a small sample size, with one SD

568 increase in the Z-standardized GRS being associated with approximately one-year earlier
569 AAD. This result is in concordance with previous results in other, larger, PD cohorts [22, 47].
570
571 Among the 90 risk variants, the *LRRK2* G2019S (rs34637584) had the largest effect estimate
572 (2.43). This variant was genotyped in our dataset but excluded during the pre-imputation QC
573 due to low MAF (0.2%). Previous studies of PD cohorts in Sweden have shown that the
574 prevalence of *LRRK2* G2019S carriers is low in Sweden, 0.54% of screened 2206 PD patients
575 were carriers [72]. It has been estimated that the G2019S mutation worldwide accounts for 4%
576 of familial and 1% of idiopathic PD cases [73]. In our study, the fraction of carriers among the
577 patients was 0.75%, further confirming a low prevalence of *LRRK2* G2019S carriers among
578 Swedish PD patients.
579
580 In conclusion, this work represents a comprehensive description of a new PD case-control
581 cohort from southern Sweden in which we nominate a novel GWA variant in the *PLPP4* locus.
582 However, subsequent studies are needed in order to validate whether *PLPP4* is associated with
583 PD within the Swedish population. This study contributes to the understanding of
584 environmental and genetic risk factors in PD in the Swedish population, and the combination
585 of epidemiological, clinical, and genetic data in the cohort makes it suitable for future studies
586 of PD etiology.
587

588 # Acknowledgments

589
590 This research was supported by MultiPark – a Strategic Research Area at Lund University and
591 by grants from the Swedish Research Council (VR), Åke Wiberg's foundation,
592 Parkinsonfonden, Åhlén foundation, Lindhés advokatbyrå, and Sigurd & Elsa Golje's

593 memorial foundation (KB/MS). This work was also supported in part by the Intramural
594 Research Program of the National Institute on Aging (NIA) (SBC/CB), part of the National
595 Institutes of Health, Department of Health and Human Services. We thank the individuals who
596 have contributed and donated blood samples to MPBC. We would also like to thank the
597 research nurses and statistician Helene Jacobsson at MultiPark for conducting data collections,
598 and previous students of the Translational Neurogenetics Unit who have helped with
599 questionnaire registration to REDCap. We would like to thank Megg Garcia and Alexander
600 Svanbergsson for technical help and feedback, and the International Parkinson's Disease
601 Genomic Consortium (IPDGC) and its trainee network for additional help during the project,
602 in particularly Lynne Krohn and Manuela Tan. We would also like to acknowledge the Global
603 Parkinson's Genetics Program (GP2) learning platform for training courses and code in
604 Parkinson's disease genetics. We would also like to acknowledge Ashfaq Ali at the National
605 Bioinformatics Infrastructure Sweden at SciLifeLab for bioinformatics advice. The Region
606 Skåne Biobank Facility (BD47) has performed the biobank services in the project.
607
608 AP receives research support from the Swedish government (ALF), Region Skåne, Skåne
609 University Hospital, Bundy Academy, Hans Gabriel and Alice Trolle Wachtmeister Stiftelse
610 för Medicinsk Forskning, The Swedish Parkinson Association, The Swedish Parkinson
611 Academy, SCA network, all in Sweden, and receives reimbursement from Elsevier Ltd. PO
612 has received research support from The Swedish National Government and County Councils,
613 Skåne University Hospital Foundations and Donations, Region Skåne, Parkinsonfonden,
614 Swedish Parkinson Academy and Åhlens Foundation.
615
616

Conflict of Interest

KB, MS, SBC, CB, AP, PO and HW have no conflicts of interest to report. OH has acquired research support (for the institution) from AVID Radiopharmaceuticals, Biogen, Eli Lilly, Eisai, GE Healthcare, Pfizer, and Roche. In the past 2 years, he has received consultancy/speaker fees from AC Immune, Alzpath, Biogen, Cerveau, and Roche.

References

[1] Collaborators GPsD (2018) Global, regional, and national burden of Parkinson's disease, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol* **17**, 939-953.

[2] Collaborators GBDN (2019) Global, regional, and national burden of neurological disorders, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol*.

[3] Blauwendraat C, Nalls MA, Singleton AB (2020) The genetic architecture of Parkinson's disease. *Lancet Neurol* **19**, 170-178.

[4] Bonifati V, Rizzu P, van Baren MJ, Schaap O, Breedveld GJ, Krieger E, Dekker MC, Squitieri F, Ibanez P, Joosse M, van Dongen JW, Vanacore N, van Swieten JC, Brice A, Meco G, van Duijn CM, Oostra BA, Heutink P (2003) Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. *Science* **299**, 256-259.

[5] Kitada T, Asakawa S, Hattori N, Matsumine H, Yamamura Y, Minoshima S, Yokochi M, Mizuno Y, Shimizu N (1998) Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature* **392**, 605-608.

641 [6] Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia A, Dutra A, Pike B, Root
642 H, Rubenstein J, Boyer R, Stenroos ES, Chandrasekharappa S, Athanassiadou A,
643 Papapetropoulos T, Johnson WG, Lazzarini AM, Duvoisin RC, Di Iorio G, Golbe LI,
644 Nussbaum RL (1997) Mutation in the alpha-synuclein gene identified in families with
645 Parkinson's disease. *Science* **276**, 2045-2047.
646 [7] Valente EM, Abou-Sleiman PM, Caputo V, Muqit MM, Harvey K, Gispert S, Ali Z,
647 Del Turco D, Bentivoglio AR, Healy DG, Albanese A, Nussbaum R, González-
648 Maldonado R, Deller T, Salvi S, Cortelli P, Gilks WP, Latchman DS, Harvey RJ,
649 Dallapiccola B, Auburger G, Wood NW (2004) Hereditary early-onset Parkinson's
650 disease caused by mutations in PINK1. *Science* **304**, 1158-1160.
651 [8] Zimprich A, Benet-Pagès A, Struhal W, Graf E, Eck SH, Offman MN, Haubenberger
652 D, Spielberger S, Schulte EC, Lichtner P, Rossle SC, Klopp N, Wolf E, Seppi K,
653 Pirker W, Presslauer S, Mollenhauer B, Katzenschlager R, Foki T, Hotzy C,
654 Reinthaler E, Harutyunyan A, Kralovics R, Peters A, Zimprich F, Brücke T, Poewe
655 W, Auff E, Trenkwalder C, Rost B, Ransmayr G, Winkelmann J, Meitinger T, Strom
656 TM (2011) A mutation in VPS35, encoding a subunit of the retromer complex, causes
657 late-onset Parkinson disease. *Am J Hum Genet* **89**, 168-175.
658 [9] Zimprich A, Biskup S, Leitner P, Lichtner P, Farrer M, Lincoln S, Kachergus J,
659 Hulihan M, Uitti RJ, Calne DB, Stoessl AJ, Pfeiffer RF, Patenge N, Carbajal IC,
660 Vieregge P, Asmus F, Müller-Myhsok B, Dickson DW, Meitinger T, Strom TM,
661 Wszolek ZK, Gasser T (2004) Mutations in LRRK2 cause autosomal-dominant
662 parkinsonism with pleomorphic pathology. *Neuron* **44**, 601-607.
663 [10] de Lau LM, Breteler MM (2006) Epidemiology of Parkinson's disease. *Lancet Neurol*
664 **5**, 525-535.

665 [11] Vollstedt EJ, Kasten M, Klein C (2019) Using global team science to identify genetic
666 parkinson's disease worldwide. *Ann Neurol* **86**, 153-157.
667 [12] Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, Tan
668 M, Kia DA, Noyce AJ, Xue A, Bras J, Young E, von Coelln R, Simón-Sánchez J,
669 Schulte C, Sharma M, Krohn L, Pihlstrøm L, Siitonen A, Iwaki H, Leonard H, Faghri
670 F, Gibbs JR, Hernandez DG, Scholz SW, Botia JA, Martinez M, Corvol J-C, Lesage
671 S, Jankovic J, Shulman LM, Sutherland M, Tienari P, Majamaa K, Toft M,
672 Andreassen OA, Bangale T, Brice A, Yang J, Gan-Or Z, Gasser T, Heutink P,
673 Shulman JM, Wood NW, Hinds DA, Hardy JA, Morris HR, Gratten J, Visscher PM,
674 Graham RR, Singleton AB, andMe Research T, System Genomics of Parkinson's
675 Disease C, International Parkinson's Disease Genomics C (2019) Identification of
676 novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-
677 analysis of genome-wide association studies. *The Lancet. Neurology* **18**, 1091-1102.
678 [13] Foo JN, Chew EGY, Chung SJ, Peng R, Blauwendraat C, Nalls MA, Mok KY, Satake
679 W, Toda T, Chao Y, Tan LCS, Tandiono M, Lian MM, Ng EY, Prakash KM, Au WL,
680 Meah WY, Mok SQ, Annuar AA, Chan AYY, Chen L, Chen Y, Jeon BS, Jiang L,
681 Lim JL, Lin JJ, Liu C, Mao C, Mok V, Pei Z, Shang HF, Shi CH, Song K, Tan AH,
682 Wu YR, Xu YM, Xu R, Yan Y, Yang J, Zhang B, Koh WP, Lim SY, Khor CC, Liu J,
683 Tan EK (2020) Identification of Risk Loci for Parkinson Disease in Asians and
684 Comparison of Risk Between Asians and Europeans: A Genome-Wide Association
685 Study. *JAMA Neurol* **77**, 746-754.
686 [14] Génin E (2020) Missing heritability of complex diseases: case solved? *Human*
687 *Genetics* **139**, 103-113.
688 [15] Young AI (2019) Solving the missing heritability problem. *PLoS Genet* **15**,
689 e1008222.

690 [16] Noyce AJ, Bestwick JP, Silveira-Moriyama L, Hawkes CH, Giovannoni G, Lees AJ,
691 Schrag A (2012) Meta-analysis of early nonmotor features and risk factors for
692 Parkinson disease. *Ann Neurol* **72**, 893-901.
693 [17] Bellou V, Belbasis L, Tzoulaki I, Evangelou E, Ioannidis JP (2016) Environmental
694 risk factors and Parkinson's disease: An umbrella review of meta-analyses.
695 *Parkinsonism Relat Disord* **23**, 1-9.
696 [18] Sirugo G, Williams SM, Tishkoff SA (2019) The Missing Diversity in Human
697 Genetic Studies. *Cell* **177**, 26-31.
698 [19] Popejoy AB, Fullerton SM (2016) Genomics is failing on diversity. *Nature* **538**, 161-
699 164.
700 [20] International Parkinson Disease Genomics C (2020) Ten Years of the International
701 Parkinson Disease Genomics Consortium: Progress and Next Steps. *Journal of*
702 *Parkinson's disease* **10**, 19-30.
703 [21] Loesch DP, Horimoto A, Heilbron K, Sarihan EI, Inca-Martinez M, Mason E,
704 Cornejo-Olivas M, Torres L, Mazzetti P, Cosentino C, Sarapura-Castro E, Rivera-
705 Valdivia A, Medina AC, Dieguez E, Raggio V, Lescano A, Tumas V, Borges V,
706 Ferraz HB, Rieder CR, Schumacher-Schuh A, Santos-Lobato BL, Velez-Pardo C,
707 Jimenez-Del-Rio M, Lopera F, Moreno S, Chana-Cuevas P, Fernandez W, Arboleda
708 G, Arboleda H, Arboleda-Bustos CE, Yearout D, Zabetian CP, Cannon P, Thornton
709 TA, O'Connor TD, Mata IF (2021) Characterizing the Genetic Architecture of
710 Parkinson's Disease in Latinos. *Ann Neurol*.
711 [22] Bandres-Ciga S, Ahmed S, Sabir MS, Blauwendraat C, Adarmes-Gómez AD, Bernal-
712 Bernal I, Bonilla-Toribio M, Buiza-Rueda D, Carrillo F, Carrión-Claro M, Gómez-
713 Garre P, Jesús S, Labrador-Espinosa MA, Macias D, Méndez-Del-Barrio C, Periñán-
714 Tocino T, Tejera-Parrado C, Vargas-González L, Diez-Fairen M, Alvarez I, Tartari

715 JP, Buongiorno M, Aguilar M, Gorostidi A, Bergareche JA, Mondragon E, Vinagre-
716 Aragon A, Croitoru I, Ruiz-Martínez J, Dols-Icardo O, Kulisevsky J, Marín-Lahoz J,
717 Pagonabarraga J, Pascual-Sedano B, Ezquerra M, Cámara A, Compta Y, Fernández
718 M, Fernández-Santiago R, Muñoz E, Tolosa E, Valldeoriola F, Gonzalez-Aramburu I,
719 Sanchez Rodriguez A, Sierra M, Menéndez-González M, Blazquez M, Garcia C,
720 Suarez-San Martin E, García-Ruiz P, Martínez-Castrillo JC, Vela-Desojo L, Ruz C,
721 Barrero FJ, Escamilla-Sevilla F, Mínguez-Castellanos A, Cerdan D, Tabernero C,
722 Gomez Heredia MJ, Perez Errazquin F, Romero-Acebal M, Feliz C, Lopez-Sendon
723 JL, Mata M, Martínez Torres I, Kim JJ, Dalgard CL, American Genome C, Brooks J,
724 Saez-Atienzar S, Gibbs JR, Jorda R, Botia JA, Bonet-Ponce L, Morrison KE, Clarke
725 C, Tan M, Morris H, Edsall C, Hernandez D, Simon-Sanchez J, Nalls MA, Scholz
726 SW, Jimenez-Escrig A, Duarte J, Vives F, Duran R, Hoenicka J, Alvarez V, Infante J,
727 Marti MJ, Clarimón J, López de Munain A, Pastor P, Mir P, Singleton A,
728 International Parkinson Disease Genomics C (2019) The Genetic Architecture of
729 Parkinson Disease in Spain: Characterizing Population-Specific Risk, Differential
730 Haplotype Structures, and Providing Etiologic Insight. *Movement disorders : official*
731 *journal of the Movement Disorder Society* **34**, 1851-1863.
732 [23] Pihlstrøm L, Axelsson G, Bjørnarå KA, Dizdar N, Fardell C, Forsgren L, Holmberg
733 B, Larsen JP, Linder J, Nissbrandt H, Tysnes OB, Ohman E, Dietrichs E, Toft M
734 (2013) Supportive evidence for 11 loci from genome-wide association studies in
735 Parkinson's disease. *Neurobiol Aging* **34**, 1708.e1707-1713.
736 [24] Humphreys K, Grankvist A, Leu M, Hall P, Liu J, Ripatti S, Rehnström K, Groop L,
737 Klareskog L, Ding B, Grönberg H, Xu J, Pedersen NL, Lichtenstein P, Mattingsdal
738 M, Andreassen OA, O'Dushlaine C, Purcell SM, Sklar P, Sullivan PF, Hultman CM,

739 Palmgren J, Magnusson PKE (2011) The genetic structure of the Swedish population.
740 *PloS one* **6**, e22547-e22547.
741 [25] Nelis M, Esko T, Mägi R, Zimprich F, Zimprich A, Toncheva D, Karachanak S,
742 Piskácková T, Balascák I, Peltonen L, Jakkula E, Rehnström K, Lathrop M, Heath S,
743 Galan P, Schreiber S, Meitinger T, Pfeufer A, Wichmann HE, Melegh B, Polgár N,
744 Toniolo D, Gasparini P, D'Adamo P, Klovins J, Nikitina-Zake L, Kucinskas V,
745 Kasnauskiene J, Lubinski J, Debniak T, Limborska S, Khrunin A, Estivill X, Rabionet
746 R, Marsal S, Julià A, Antonarakis SE, Deutsch S, Borel C, Attar H, Gagnebin M,
747 Macek M, Krawczak M, Remm M, Metspalu A (2009) Genetic structure of
748 Europeans: a view from the North-East. *PloS one* **4**, e5472-e5472.
749 [26] Ameur A, Dahlberg J, Olason P, Vezzi F, Karlsson R, Martin M, Viklund J, Kähäri
750 AK, Lundin P, Che H, Thutkawkorapin J, Eisfeldt J, Lampa S, Dahlberg M, Hagberg
751 J, Jareborg N, Liljedahl U, Jonasson I, Johansson Å, Feuk L, Lundeberg J, Syvänen
752 AC, Lundin S, Nilsson D, Nystedt B, Magnusson PK, Gyllensten U (2017) SweGen:
753 a whole-genome data resource of genetic variability in a cross-section of the Swedish
754 population. *Eur J Hum Genet* **25**, 1253-1260.
755 [27] Salmela E, Lappalainen T, Liu J, Sistonen P, Andersen PM, Schreiber S, Savontaus
756 M-L, Czene K, Lahermo P, Hall P, Kere J (2011) Swedish population substructure
757 revealed by genome-wide single nucleotide polymorphism data. *PloS one* **6**, e16747-
758 e16747.
759 [28] Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG (2009) Research
760 electronic data capture (REDCap)—A metadata-driven methodology and workflow
761 process for providing translational research informatics support. *Journal of*
762 *Biomedical Informatics* **42**, 377-381.
763 [29] R Core Team (2020) R Foundation for Statistical Computing, Vienna, Austria.

764 [30] Burström K, Sun S, Gerdtham UG, Henriksson M, Johannesson M, Levin L,
765 Zethraeus N (2014) Swedish experience-based value sets for EQ-5D health states.
766 *Qual Life Res* **23**, 431-442.
767 [31] Burström K, Teni FS, Gerdtham UG, Leidl R, Helgesson G, Rolfson O, Henriksson
768 M (2020) Experience-Based Swedish TTO and VAS Value Sets for EQ-5D-5L Health
769 States. *Pharmacoeconomics* **38**, 839-856.
770 [32] Textor J, van der Zander B, Gilthorpe MS, Liśkiewicz M, Ellison GTH (2016) Robust
771 causal inference using directed acyclic graphs: the R package 'dagitty'. *International*
772 *Journal of Epidemiology* **45**, 1887-1894.
773 [33] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J,
774 Sklar P, de Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-
775 genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-
776 575.
777 [34] Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015) Second-
778 generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*
779 **4**, 7.
780 [35] Gibbs RA, Belmont JW, Hardenbol P, Willis TD, Yu F, Yang H, Ch'ang L-Y, Huang
781 W, Liu B, Shen Y, Tam PK-H, Tsui L-C, Waye MMY, Wong JT-F, Zeng C, Zhang
782 Q, Chee MS, Galver LM, Kruglyak S, Murray SS, Oliphant AR, Montpetit A, Hudson
783 TJ, Chagnon F, Ferretti V, Leboeuf M, Phillips MS, Verner A, Kwok P-Y, Duan S,
784 Lind DL, Miller RD, Rice JP, Saccone NL, Taillon-Miller P, Xiao M, Nakamura Y,
785 Sekine A, Sorimachi K, Tanaka T, Tanaka Y, Tsunoda T, Yoshino E, Bentley DR,
786 Deloukas P, Hunt S, Powell D, Altshuler D, Gabriel SB, Zhang H, Zeng C, Matsuda I,
787 Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Aniagwu T,
788 Marshall PA, Matthew O, Nkwodimmah C, Royal CDM, Leppert MF, Dixon M,

789 Stein LD, Cunningham F, Kanani A, Thorisson GA, Chakravarti A, Chen PE, Cutler
790 DJ, Kashuk CS, Donnelly P, Marchini J, McVean GAT, Myers SR, Cardon LR,
791 Abecasis GR, Morris A, Weir BS, Mullikin JC, Sherry ST, Feolo M, Altshuler D,
792 Daly MJ, Schaffner SF, Qiu R, Kent A, Dunston GM, Kato K, Niikawa N, Knoppers
793 BM, Foster MW, Clayton EW, Wang VO, Watkin J, Gibbs RA, Belmont JW,
794 Sodergren E, Weinstock GM, Wilson RK, Fulton LL, Rogers J, Birren BW, Han H,
795 Wang H, Godbout M, Wallenburg JC, L'Archevêque P, Bellemare G, Todani K,
796 Fujita T, Tanaka S, Holden AL, Lai EH, Collins FS, Brooks LD, McEwen JE, Guyer
797 MS, Jordan E, Peterson JL, Spiegel J, Sung LM, Zacharia LF, Kennedy K, Dunn MG,
798 Seabrook R, Shillito M, Skene B, Stewart JG, Valle DL, Clayton EW, Jorde LB,
799 Belmont JW, Chakravarti A, Cho MK, Duster T, Foster MW, Jasperse M, Knoppers
800 BM, Kwok P-Y, Licinio J, Long JC, Marshall PA, Ossorio PN, Wang VO, Rotimi
801 CN, Royal CDM, Spallone P, Terry SF, Lander ES, Lai EH, Nickerson DA, Abecasis
802 GR, Altshuler D, Bentley DR, Boehnke M, Cardon LR, Daly MJ, Deloukas P,
803 Douglas JA, Gabriel SB, Hudson RR, Hudson TJ, Kruglyak L, Kwok P-Y, Nakamura
804 Y, Nussbaum RL, Royal CDM, Schaffner SF, Sherry ST, Stein LD, Tanaka T, †The
805 International HapMap C, Genotyping centres: Baylor College of M, ParAllele B,
806 Chinese HapMap C, Illumina, McGill U, Génome Québec Innovation C, University
807 of California at San F, Washington U, University of T, Riken, Wellcome Trust Sanger
808 I, Whitehead Institute MITCfGR, Community engagement/public c, sample-
809 collection groups: Beijing Normal U, Beijing Genomics I, Health Sciences University
810 of Hokkaido EEI, Shinshu U, Howard U, University of I, University of U, Analysis
811 Groups: Cold Spring Harbor L, Johns Hopkins University School of M, University of
812 O, University of Oxford WTCfHG, Health USNIo, Ethical L, Social Issues: Chinese
813 Academy of Social S, Genetic Interest G, Howard U, Kyoto U, Nagasaki U,

814 University of M, University of O, Vanderbilt U, Wellcome T, Medicine SNPDBCo,
815 Washington U, Scientific Management: Chinese Academy of S, Chinese Ministry of
816 S, Technology, Genome C, Génome Q, Japanese Ministry of Education CSS,
817 Technology, The SNPC, Initial Planning Groups P, Ethical L, Social Issues G,
818 Methods G (2003) The International HapMap Project. *Nature* **426**, 789-796.
819 [36] Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, Vrieze SI, Chew EY,
820 Levy S, McGue M, Schlessinger D, Stambolian D, Loh PR, Iacono WG, Swaroop A,
821 Scott LJ, Cucca F, Kronenberg F, Boehnke M, Abecasis GR, Fuchsberger C (2016)
822 Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-1287.
823 [37] Loh PR, Danecek P, Palamara PF, Fuchsberger C, Y AR, H KF, Schoenherr S, Forer
824 L, McCarthy S, Abecasis GR, Durbin R, A LP (2016) Reference-based phasing using
825 the Haplotype Reference Consortium panel. *Nat Genet* **48**, 1443-1448.
826 [38] McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, Kang HM,
827 Fuchsberger C, Danecek P, Sharp K, Luo Y, Sidore C, Kwong A, Timpson N,
828 Koskinen S, Vrieze S, Scott LJ, Zhang H, Mahajan A, Veldink J, Peters U, Pato C,
829 van Duijn CM, Gillies CE, Gandin I, Mezzavilla M, Gilly A, Cocca M, Traglia M,
830 Angius A, Barrett JC, Boomsma D, Branham K, Breen G, Brummett CM, Busonero
831 F, Campbell H, Chan A, Chen S, Chew E, Collins FS, Corbin LJ, Smith GD,
832 Dedoussis G, Dorr M, Farmaki AE, Ferrucci L, Forer L, Fraser RM, Gabriel S, Levy
833 S, Groop L, Harrison T, Hattersley A, Holmen OL, Hveem K, Kretzler M, Lee JC,
834 McGue M, Meitinger T, Melzer D, Min JL, Mohlke KL, Vincent JB, Nauck M,
835 Nickerson D, Palotie A, Pato M, Pirastu N, McInnis M, Richards JB, Sala C, Salomaa
836 V, Schlessinger D, Schoenherr S, Slagboom PE, Small K, Spector T, Stambolian D,
837 Tuke M, Tuomilehto J, Van den Berg LH, Van Rheenen W, Volker U, Wijmenga C,
838 Toniolo D, Zeggini E, Gasparini P, Sampson MG, Wilson JF, Frayling T, de Bakker

839 PI, Swertz MA, McCarroll S, Kooperberg C, Dekker A, Altshuler D, Willer C, Iacono
840 W, Ripatti S, Soranzo N, Walter K, Swaroop A, Cucca F, Anderson CA, Myers RM,
841 Boehnke M, McCarthy MI, Durbin R (2016) A reference panel of 64,976 haplotypes
842 for genotype imputation. *Nat Genet* **48**, 1279-1283.
843 [39] Zhan X, Hu Y, Li B, Abecasis GR, Liu DJ (2016) RVTESTS: an efficient and
844 comprehensive tool for rare variant association analysis using sequence data.
845 *Bioinformatics* **32**, 1423-1426.
846 [40] Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic
847 variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164.
848 [41] Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J,
849 DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper
850 R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype
851 blocks in the human genome. *Science* **296**, 2225-2229.
852 [42] Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of
853 LD and haplotype maps. *Bioinformatics* **21**, 263-265.
854 [43] Uhl GR, Drgon T, Johnson C, Walther D, David SP, Aveyard P, Murphy M,
855 Johnstone EC, Munafò MR (2010) Genome-wide association for smoking cessation
856 success: participants in the Patch in Practice trial of nicotine replacement.
857 *Pharmacogenomics* **11**, 357-367.
858 [44] Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of
859 LD and haplotype maps. *Bioinformatics* **21**, 263-265.
860 [45] Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, Christiani
861 DC, Wurfel MM, Lin X (2012) Optimal unified approach for rare-variant association
862 testing with application to small-sample case-control whole-exome sequencing
863 studies. *Am J Hum Genet* **91**, 224-237.

[46] Nalls MA, McLean CY, Rick J, Eberly S, Hutten SJ, Gwinn K, Sutherland M, Martinez M, Heutink P, Williams NM, Hardy J, Gasser T, Brice A, Price TR, Nicolas A, Keller MF, Molony C, Gibbs JR, Chen-Plotkin A, Suh E, Letson C, Fiandaca MS, Mapstone M, Federoff HJ, Noyce AJ, Morris H, Van Deerlin VM, Weintraub D, Zabetian C, Hernandez DG, Lesage S, Mullins M, Conley ED, Northover CAM, Frasier M, Marek K, Day-Williams AG, Stone DJ, Ioannidis JPA, Singleton AB, Parkinson's Disease Biomarkers P, Parkinson's Progression Marker Initiative i (2015) Diagnosis of Parkinson's disease on the basis of clinical and genetic classification: a population-based modelling study. *The Lancet. Neurology* **14**, 1002-1009.

[47] Blauwendraat C, Heilbron K, Vallerga CL, Bandres-Ciga S, von Coelln R, Pihlstrom L, Simon-Sanchez J, Schulte C, Sharma M, Krohn L, Siitonen A, Iwaki H, Leonard H, Noyce AJ, Tan M, Gibbs JR, Hernandez DG, Scholz SW, Jankovic J, Shulman LM, Lesage S, Corvol JC, Brice A, van Hilten JJ, Marinus J, Eerola-Rautio J, Tienari P, Majamaa K, Toft M, Grosset DG, Gasser T, Heutink P, Shulman JM, Wood N, Hardy J, Morris HR, Hinds DA, Gratten J, Visscher PM, Gan-Or Z, Nalls MA, Singleton AB (2019) Parkinson's disease age at onset genome-wide association study: Defining heritability, genetic loci, and alpha-synuclein mechanisms. *Mov Disord*.

[48] Taliun D, Harris DN, Kessler MD, Carlson J, Szpiech ZA, Torres R, Taliun SAG, Corvelo A, Gogarten SM, Kang HM, Pitsillides AN, LeFaive J, Lee S-b, Tian X, Browning BL, Das S, Emde A-K, Clarke WE, Loesch DP, Shetty AC, Blackwell TW, Smith AV, Wong Q, Liu X, Conomos MP, Bobo DM, Aguet F, Albert C, Alonso A, Ardlie KG, Arking DE, Aslibekyan S, Auer PL, Barnard J, Barr RG, Barwick L, Becker LC, Beer RL, Benjamin EJ, Bielak LF, Blangero J, Boehnke M, Bowden DW, Brody JA, Burchard EG, Cade BE, Casella JF, Chalazan B, Chasman DI, Chen Y-DI, Cho MH, Choi SH, Chung MK, Clish CB, Correa A, Curran JE, Custer B, Darbar D,

Daya M, de Andrade M, DeMeo DL, Dutcher SK, Ellinor PT, Emery LS, Eng C, Fatkin D, Fingerlin T, Forer L, Fornage M, Franceschini N, Fuchsberger C, Fullerton SM, Germer S, Gladwin MT, Gottlieb DJ, Guo X, Hall ME, He J, Heard-Costa NL, Heckbert SR, Irvin MR, Johnsen JM, Johnson AD, Kaplan R, Kardia SLR, Kelly T, Kelly S, Kenny EE, Kiel DP, Klemmer R, Konkle BA, Kooperberg C, Köttgen A, Lange LA, Lasky-Su J, Levy D, Lin X, Lin K-H, Liu C, Loos RJF, Garman L, Gerszten R, Lubitz SA, Lunetta KL, Mak ACY, Manichaikul A, Manning AK, Mathias RA, McManus DD, McGarvey ST, Meigs JB, Meyers DA, Mikulla JL, Minear MA, Mitchell BD, Mohanty S, Montasser ME, Montgomery C, Morrison AC, Murabito JM, Natale A, Natarajan P, Nelson SC, North KE, O'Connell JR, Palmer ND, Pankratz N, Peloso GM, Peyser PA, Pleiness J, Post WS, Psaty BM, Rao DC, Redline S, Reiner AP, Roden D, Rotter JI, Ruczinski I, Sarnowski C, Schoenherr S, Schwartz DA, Seo J-S, Seshadri S, Sheehan VA, Sheu WH, Shoemaker MB, Smith NL, Smith JA, Sotoodehnia N, Stilp AM, Tang W, Taylor KD, Telen M, Thornton TA, Tracy RP, Van Den Berg DJ, Vasan RS, Viaud-Martinez KA, Vrieze S, Weeks DE, Weir BS, Weiss ST, Weng L-C, Willer CJ, Zhang Y, Zhao X, Arnett DK, Ashley-Koch AE, Barnes KC, Boerwinkle E, Gabriel S, Gibbs R, Rice KM, Rich SS, Silverman EK, Qasba P, Gan W, Abe N, Almasy L, Ament S, Anderson P, Anugu P, Applebaum-Bowden D, Assimes T, Avramopoulos D, Barron-Casella E, Beaty T, Beck G, Becker D, Beitelshees A, Benos T, Bezerra M, Bis J, Bowler R, Broeckel U, Broome J, Bunting K, Bustamante C, Buth E, Cardwell J, Carey V, Carty C, Casaburi R, Castaldi P, Chaffin M, Chang C, Chang Y-C, Chavan S, Chen B-J, Chen W-M, Chuang L-M, Chung R-H, Comhair S, Cornell E, Crandall C, Crapo J, Curtis J, Damcott C, David S, Davis C, Fuentes Ldl, DeBaun M, Deka R, Devine S, Duan Q, Duggirala R, Durda JP, Eaton C, Ekunwe L, El Boueiz A, Erzurum S, Farber C,

914 Flickinger M, Fornage M, Frazar C, Fu M, Fulton L, Gao S, Gao Y, Gass M, Gelb B,
915 Geng XP, Geraci M, Ghosh A, Gignoux C, Glahn D, Gong D-W, Goring H, Graw S,
916 Grine D, Gu CC, Guan Y, Gupta N, Haessler J, Hawley NL, Heavner B, Herrington
917 D, Hersh C, Hidalgo B, Hixson J, Hobbs B, Hokanson J, Hong E, Hoth K, Hsiung
918 CA, Hung Y-J, Huston H, Hwu CM, Jackson R, Jain D, Jhun MA, Johnson C,
919 Johnston R, Jones K, Kathiresan S, Khan A, Kim W, Kinney G, Kramer H, Lange C,
920 Lange E, Lange L, Laurie C, LeBoff M, Lee J, Lee SS, Lee W-J, Levine D, Lewis J,
921 Li X, Li Y, Lin H, Lin H, Lin KH, Liu S, Liu Y, Liu Y, Luo J, Mahaney M,
922 Consortium NT-OfPM (2021) Sequencing of 53,831 diverse genomes from the
923 NHLBI TOPMed Program. *Nature* **590**, 290-299.
924 [49] Schapira AHV, Chaudhuri KR, Jenner P (2017) Non-motor features of Parkinson
925 disease. *Nature Reviews Neuroscience* **18**, 435-450.
926 [50] Kashihara K (2006) Weight loss in Parkinson's disease. *J Neurol* **253 Suppl 7**, Vii38-
927 41.
928 [51] Pandey S, Srivanitchapoom P (2017) Levodopa-induced Dyskinesia: Clinical
929 Features, Pathophysiology, and Medical Management. *Annals of Indian Academy of*
930 *Neurology* **20**, 190-198.
931 [52] Freire C, Koifman S (2012) Pesticide exposure and Parkinson's disease:
932 Epidemiological evidence of association. *NeuroToxicology* **33**, 947-971.
933 [53] KEMI, Statistics on pesticides, Swedish Chemicals Agency,
934 https://www.kemi.se/en/statistics-on-chemicals-in-sweden/statistics-on-pesticides,
935 November 27 2020, Accessed May 5 2021.
936 [54] Sellbach AN, Boyle RS, Silburn PA, Mellick GD (2006) Parkinson's disease and
937 family history. *Parkinsonism & Related Disorders* **12**, 399-409.

938 [55] Marras C, Hincapié CA, Kristman VL, Cancelliere C, Soklaridis S, Li A, Borg J, af
939 Geijerstam JL, Cassidy JD (2014) Systematic review of the risk of Parkinson's disease
940 after mild traumatic brain injury: results of the International Collaboration on Mild
941 Traumatic Brain Injury Prognosis. *Arch Phys Med Rehabil* **95**, S238-244.
942 [56] Jafari S, Etminan M, Aminzadeh F, Samii A (2013) Head injury and risk of Parkinson
943 disease: a systematic review and meta-analysis. *Mov Disord* **28**, 1222-1229.
944 [57] Holm H, Jarvis MJ, Russell MA, Feyerabend C (1992) Nicotine intake and
945 dependence in Swedish snuff takers. *Psychopharmacology (Berl)* **108**, 507-511.
946 [58] Yang F, Pedersen NL, Ye W, Liu Z, Norberg M, Forsgren L, Trolle Lagerros Y,
947 Bellocco R, Alfredsson L, Knutsson A, Jansson JH, Wennberg P, Galanti MR, Lager
948 ACJ, Araghi M, Lundberg M, Magnusson C, Wirdefeldt K (2017) Moist smokeless
949 tobacco (Snus) use and risk of Parkinson's disease. *Int J Epidemiol* **46**, 872-880.
950 [59] Tobacco, Public Health Agency of Sweden,
951 https://www.folkhalsomyndigheten.se/the-public-health-agency-of-sweden/living-
952 conditions-and-lifestyle/alcohol-narcotics-doping-tobacco-and-gambling/tobacco/,
953 June 28 2019, Accessed April 20.
954 [60] Quik M (2004) Smoking, nicotine and Parkinson's disease. *Trends in Neurosciences*
955 **27**, 561-568.
956 [61] Hou L, Li Q, Jiang L, Qiu H, Geng C, Hong J-S, Li H, Wang Q (2018) Hypertension
957 and Diagnosis of Parkinson's Disease: A Meta-Analysis of Cohort Studies. *Frontiers*
958 *in neurology* **9**, 162-162.
959 [62] Jiang Z, Xu X, Gu X, Ou R, Luo X, Shang H, Song W (2020) Effects of Higher
960 Serum Lipid Levels on the Risk of Parkinson's Disease: A Systematic Review and
961 Meta-Analysis. *Front Neurol* **11**, 597.

962 [63] Yan J, Qiao L, Tian J, Liu A, Wu J, Huang J, Shen M, Lai X (2019) Effect of statins
963 on Parkinson's disease: A systematic review and meta-analysis. *Medicine* **98**.
964 [64] Jones JD, Malaty I, Price CC, Okun MS, Bowers D (2012) Health comorbidities and
965 cognition in 1948 patients with idiopathic Parkinson's disease. *Parkinsonism Relat*
966 *Disord* **18**, 1073-1078.
967 [65] Rugbjerg K, Friis S, Jørgensen TL, Ritz B, Korbo L, Olsen JH (2010) Risk for
968 Parkinson's disease among patients with osteoarthritis: a Danish cohort study. *Mov*
969 *Disord* **25**, 2355-2360.
970 [66] Gao X, Chen H, Schwarzschild MA, Ascherio A (2011) Use of ibuprofen and risk of
971 Parkinson disease. *Neurology* **76**, 863-869.
972 [67] Brakedal B, Tzoulis C, Tysnes OB, Haugarvoll K (2021) NSAID use is not associated
973 with Parkinson's disease incidence: A Norwegian Prescription Database study. *PLoS*
974 *One* **16**, e0256602.
975 [68] Zhang X, Zhang L, Lin B, Chai X, Li R, Liao Y, Deng X, Liu Q, Yang W, Cai Y,
976 Zhou W, Lin Z, Huang W, Zhong M, Lei F, Wu J, Yu S, Li X, Li S, Li Y, Zeng J,
977 Long W, Ren D, Huang Y (2017) Phospholipid Phosphatase 4 promotes proliferation
978 and tumorigenesis, and activates Ca(2+)-permeable Cationic Channel in lung
979 carcinoma cells. *Mol Cancer* **16**, 147.
980 [69] Simón-Sánchez J, van Hilten JJ, van de Warrenburg B, Post B, Berendse HW,
981 Arepalli S, Hernandez DG, de Bie RM, Velseboer D, Scheffer H, Bloem B, van Dijk
982 KD, Rivadeneira F, Hofman A, Uitterlinden AG, Rizzu P, Bochdanovits Z, Singleton
983 AB, Heutink P (2011) Genome-wide association study confirms extant PD risk loci
984 among the Dutch. *Eur J Hum Genet* **19**, 655-661.
985 [70] Nalls MA, Pankratz N, Lill CM, Do CB, Hernandez DG, Saad M, DeStefano AL,
986 Kara E, Bras J, Sharma M, Schulte C, Keller MF, Arepalli S, Letson C, Edsall C,

987 Stefansson H, Liu X, Pliner H, Lee JH, Cheng R, International Parkinson's Disease
988 Genomics C, Parkinson's Study Group Parkinson's Research: The Organized GI,
989 andMe, GenePd, NeuroGenetics Research C, Hussman Institute of Human G,
990 Ashkenazi Jewish Dataset I, Cohorts for H, Aging Research in Genetic E, North
991 American Brain Expression C, United Kingdom Brain Expression C, Greek
992 Parkinson's Disease C, Alzheimer Genetic Analysis G, Ikram MA, Ioannidis JPA,
993 Hadjigeorgiou GM, Bis JC, Martinez M, Perlmutter JS, Goate A, Marder K, Fiske B,
994 Sutherland M, Xiromerisiou G, Myers RH, Clark LN, Stefansson K, Hardy JA,
995 Heutink P, Chen H, Wood NW, Houlden H, Payami H, Brice A, Scott WK, Gasser T,
996 Bertram L, Eriksson N, Foroud T, Singleton AB (2014) Large-scale meta-analysis of
997 genome-wide association data identifies six new risk loci for Parkinson's disease.
998 *Nature genetics* **46**, 989-993.
999 [71] Fanning S, Selkoe D, Dettmer U (2020) Parkinson's disease: proteinopathy or
1000 lipidopathy? *npj Parkinson's Disease* **6**, 3.
1001 [72] Puschmann A, Jiménez-Ferrer I, Lundblad-Andersson E, Mårtensson E, Hansson O,
1002 Odin P, Widner H, Brolin K, Mzezewa R, Kristensen J, Soller M, Rödström EY, Ross
1003 OA, Toft M, Breedveld GJ, Bonifati V, Brodin L, Zettergren A, Sydow O, Linder J,
1004 Wirdefeldt K, Svenningsson P, Nissbrandt H, Belin AC, Forsgren L, Swanberg M
1005 (2019) Low prevalence of known pathogenic mutations in dominant PD genes: A
1006 Swedish multicenter study. *Parkinsonism & related disorders* **66**, 158-165.
1007 [73] Tolosa E, Vila M, Klein C, Rascol O (2020) LRRK2 in Parkinson disease: challenges
1008 of clinical trials. *Nature Reviews Neurology* **16**, 97-107.
1009
1010

Tables

Table 1: Characteristics of participants (individuals with Parkinson's disease (PD) and individuals without the disease [controls]) in the MPBC cohort.

	Patients (N=929)	Controls (N=935)
Sex		
n	929	935
Men	599 (64.5%)	598 (64.0%)
Women	330 (35.5%)	337 (36.0%)
Birth year		
n	929	935
Mean (SD)	1944.4 (8.4)	1944.4 (8.3)
Range	1920.0-1979.0	1920.0-1980.0
Age at inclusion		
n	929	935
Mean (SD)	71.0 (8.3)	72.3 (8.3)
Range	37.0-96.0	38.0-97.0
Swedish ancestry		
n	922	921
No	112 (12.1%)	128 (13.9%)
Yes	810 (87.9%)	793 (86.1%)
Highest completed education		
n	923	929
< Primary and lower secondary education	8 (0.9%)	4 (0.4%)
Primary and lower secondary education	241 (26.1%)	232 (25.0%)
Upper secondary education	403 (43.7%)	384 (41.3%)
University	271 (29.4%)	309 (33.3%)
BMI at inclusion (kg/m^2)		
n	914	921
Mean (SD)	25.8 (4.4)	26.0 (3.9)
Range	15.2-48.8	15.9-42.3
Marital status		
n	925	931
Single	57 (6.2%)	62 (6.7%)
Married/Civil partnership	719 (77.7%)	713 (76.6%)
Divorced/Separated	71 (7.7%)	60 (6.4%)
Widow/Widower	78 (8.4%)	96 (10.3%)
TTO Index		
n	898	903
Mean (SD)	0.8 (0.1)	0.9 (0.1)
Range	0.4-1.0	0.5-1.0
VAS Index		
n	898	903
Mean (SD)	68.9 (16.8)	82.0 (9.7)
Range	17.2-88.9	20.9-88.9

Swedish ancestry: Both parents were born in Sweden. BMI: Body Mass Index; TTO: Time Trade Of, EQ-5D-3L; VAS: Visual Analogue Scale, EQ-5D-3L

Table 2: Characteristics of individuals with Parkinson's disease in the MPBC cohort.

	Men (N=599)	Women (N=330)
Age at diagnosis		
n	503	289
Median (Q1, Q3)	67.0 (60.0, 72.0)	67.0 (59.0, 73.0)
Range	29.0-84.0	35.0-89.0
Disease length at inclusion		
n	503	289
Median (Q1, Q3)	4.0 (1.5, 8.0)	4.0 (2.0, 7.0)
Range	0.0-36.0	0.0-33.0
Hoehn & Yahr		
n	514	295
Median (Q1, Q3)	2.0 (1.5, 3.0)	2.0 (1.5, 3.0)
Range	0.0-5.0	0.0-5.0
CISI-PD		
n	518	301
Median (Q1, Q3)	5.0 (3.0, 8.0)	5.0 (2.0, 7.0)
Range	0.0-21.0	0.0-20.0
PDQ-8		
n	518	299
Median (Q1, Q3)	6.0 (2.0, 10.0)	7.0 (3.0, 11.0)
Range	0.0-25.0	0.0-21.0

Data retrieved from the Swedish Parkinson Registry. CISI-PD: Clinical Impression of Severity Index – Parkinson's Disease, summary score; PDQ-8: Parkinson's Disease Questionnaire – 8, summary score.

Table 3: Identified haplotypes in the *PLPP4* locus

Locus	SNPs (N)	Haplotype	MAF Pat.	MAF Ctrls	KB	OR[a]	95%CI[a]	P-value[a]	OR[b]	95%CI[b]	P-value[b]	OR[c]	95%CI	P-value[c]	SNPs
H1	8	GACGGGGG	0.901	0.911	41.334	1.00	Ref.	3.00E-01	1.00	Ref.	3.20E-01	1.00	Ref.	3.00E-01	rs2420703|rs71484626|rs2420706|rs2420707|rs2901235|rs12571802|rs2901237|rs17100130
H1	8	AGAAAAAA	0.013	0.012	41.334	1.19	0.92-1.5	1.73E-01	1.19	0.92-1.5	8.20E-01	1.20	0.92-1.6	1.70E-01	rs2420703|rs71484626|rs2420706|rs2420707|rs2901235|rs12571802|rs2901237|rs17100130
H1	8	GGCGAGGG	0.076	0.065	41.334	1.09	0.60-2.0	8.10E-01	1.08	0.59-2.0	1.80E-01	1.07	0.58-1.9	8.60E-01	rs2420703|rs71484626|rs2420706|rs2420707|rs2901235|rs12571802|rs2901237|rs17100130
H2	2	GG	0.750	0.763	10.817	1.00	Ref.	3.28E-01	1.00	Ref.	3.16E-01	1.00	Ref.	4.39E-01	rs7068076|rs36108282
H2	2	AA	0.071	0.047	10.817	1.56	1.17-2.06	1.48E-03	1.54	1.16-2.04	2.31E-03	1.48	1.11-1.97	5.57E-03	rs7068076|rs36108282
H2	2	AG	0.178	0.188	10.817	0.97	0.82-1.14	4.15E-01	0.97	0.82-1.15	4.74E-01	0.96	0.81-1.14	4.23E-01	rs7068076|rs36108282
H3	3	GAG	0.448	0.450	7.153	1.00	Ref.	9.07E-01	1.00	Ref.	9.93E-01	1.00	Ref.	8.53E-01	rs4595466|rs2244544|rs77772193
H3	3	GGG	0.129	0.094	7.153	1.34	1.08-1.65	1.30E-03	1.33	1.07-1.65	1.36E-03	1.31	1.06-1.63	1.99E-03	rs4595466|rs2244544|rs77772193
H3	3	GGA	0.082	0.070	7.153	1.19	0.92-1.53	1.56E-01	1.17	0.91-1.51	1.83E-01	1.14	0.89-1.48	2.38E-01	rs4595466|rs2244544|rs77772193
H3	3	AGG	0.341	0.386	7.153	0.88	0.76-1.02	4.12E-03	0.88	0.76-1.01	3.30E-03	0.87	0.75-1.01	3.10E-03	rs4595466|rs2244544|rs77772193
H4	2	CA	0.492	0.512	5.468	1.00	Ref.	2.35E-01	1.00	Ref.	2.71E-01	1.00	Ref.	3.12E-01	rs60288721|rs1536389
H4	2	CG	0.432	0.415	5.468	1.08	0.95-1.24	2.83E-01	1.07	0.94-1.23	3.44E-01	1.07	0.94-1.22	3.65E-01	rs60288721|rs1536389
H4	2	AA	0.075	0.073	5.468	1.07	0.83-1.38	8.08E-01	1.08	0.83-1.39	7.49E-01	1.06	0.82-1.37	8.24E-01	rs60288721|rs1536389
H5	2	CA	0.482	0.494	5.908	1.00	Ref.	4.72E-01	1.00	Ref.	5.06E-01	1.00	Ref.	5.46E-01	rs11199378|rs12768425
H5	2	AC	0.491	0.477	5.908	1.06	0.93-1.06	3.82E-01	1.05	0.92-1.20	4.33E-01	1.04	0.92-1.19	5.04E-01	rs11199378|rs12768425
H5	2	AA	0.011	0.009	5.908	1.21	0.63-2.32	6.15E-01	1.19	0.61-2.30	6.51E-01	1.26	0.65-2.45	5.28E-01	rs11199378|rs12768425
H5	2	CC	0.016	0.020	5.908	0.81	0.50-1.32	3.34E-01	0.84	0.51-1.39	4.30E-01	0.86	0.52-1.41	4.75E-01	rs11199378|rs12768425
H6	5	GGAGA	0.408	0.359	41.913	1.00	Ref.	2.03E-03	1.00	Ref.	1.59E-03	1.00	Ref.	1.59E-03	rs978854|rs7910507|rs10886711|rs11199417|rs10886717
H6	5	GAGGA	0.039	0.028	41.913	1.17	0.82-1.68	1.03E-01	1.19	0.83-1.70	8.96E-02	1.15	0.80-1.65	1.29E-01	rs978854|rs7910507|rs10886711|rs11199417|rs10886717
H6	5	AGGAA	0.209	0.189	41.913	0.97	0.81-1.16	1.16E-01	0.98	0.82-1.17	9.48E-02	0.97	0.81-1.16	1.08E-01	rs978854|rs7910507|rs10886711|rs11199417|rs10886717
H6	5	GAGGG	0.050	0.050	41.913	0.91	0.67-1.24	8.35E-01	0.90	0.66-1.24	8.67E-01	0.91	0.66-1.25	8.08E-01	rs978854|rs7910507|rs10886711|rs11199417|rs10886717
H6	**5**	**GGGGA**	**0.242**	**0.302**	**41.913**	**0.70**	**0.60-0.83**	**4.67E-05**	**0.70**	**0.59-0.82**	**3.28E-05**	**0.69**	**0.59-0.82**	**2.54E-05**	**rs978854|rs7910507|rs10886711|rs11199417|rs10886717**
H6	5	AGGGA	0.045	0.063	41.913	0.59	0.44-0.81	7.14E-03	0.59	0.43-0.81	7.29E-03	0.97	0.81-1.16	8.92E-03	rs978854|rs7910507|rs10886711|rs11199417|rs10886717
H7	3	GGG	0.439	0.408	2.751	1.00	Ref.	6.06E-02	1.00	Ref.	9.14E-02	1.00	Ref.	7.22E-02	rs75318356|rs7912073|rs11199429
H7	3	GGA	0.403	0.402	2.751	0.93	0.81-1.07	9.55E-01	0.94	0.82-1.09	8.91E-01	0.93	0.81-1.08	9.83E-01	rs75318356|rs7912073|rs11199429
H7	3	AAA	0.115	0.123	2.751	0.87	0.70-1.07	3.95E-01	0.88	0.71-1.10	4.86E-01	0.87	0.70-1.08	4.20E-01	rs75318356|rs7912073|rs11199429
H7	3	GAA	0.044	0.065	2.751	0.63	0.47-0.85	4.83E-03	0.63	0.46-0.85	5.00E-03	0.64	0.47-0.87	7.71E-03	rs75318356|rs7912073|rs11199429
H8	2	GG	0.663	0.683	0.417	1.00	Ref.	1.85E-01	1.00	Ref.	2.16E-01	1.00	Ref.	2.09E-01	rs9943454|rs41287958
H8	2	AG	0.259	0.250	0.417	1.07	0.92-1.24	5.11E-01	1.07	0.92-1.24	5.14E-01	1.07	0.92-1.25	4.71E-01	rs9943454|rs41287958
H8	2	GA	0.078	0.067	0.417	1.21	0.93-1.56	1.95E-01	1.18	0.91-1.53	2.54E-01	1.17	0.90-1.52	2.91E-01	rs9943454|rs41287958
H9	2	AG	0.523	0.512	6.032	1.00	Ref.	5.01E-01	1.00	Ref.	5.12E-01	1.00	Ref.	5.39E-01	rs2420779|rs12415215
H9	2	CG	0.428	0.436	6.032	0.96	0.84-1.10	6.20E-01	0.96	0.84-1.10	6.28E-01	0.96	0.84-1.10	6.18E-01	rs2420779|rs12415215
H9	2	AA	0.050	0.052	6.032	0.92	0.69-1.24	6.90E-01	0.93	0.69-1.25	6.97E-01	0.94	0.70-1.28	7.91E-01	rs2420779|rs12415215
H10	3	AGA	0.291	0.293	12.005	1.00	Ref.	9.01E-01	1.00	Ref.	9.89E-01	1.00	Ref.	9.86E-01	rs10886725|rs7068008|rs4752444
H10	3	AGG	0.273	0.248	12.005	1.11	0.93-1.41	9.20E-02	1.11	0.93-1.32	7.26E-02	1.11	0.93-1.32	6.41E-02	rs10886725|rs7068008|rs4752444
H10	3	AAG	0.280	0.280	12.005	1.01	0.85-1.20	9.95E-01	0.99	0.84-1.18	8.87E-01	1.00	0.84-1.19	9.09E-01	rs10886725|rs7068008|rs4752444
H10	3	GAG	0.156	0.178	12.005	0.89	0.73-1.08	7.51E-02	0.87	0.72-10.6	5.85E-02	0.87	0.71-1.06	4.66E-02	rs10886725|rs7068008|rs4752444
H11	3	AGG	0.523	0.533	0.866	1.00	Ref.	5.42E-01	1.00	Ref.	5.91E-01	1.00	Ref.	6.15E-01	rs2420768|rs7914572|rs11594222
H11	3	GAG	0.249	0.235	0.866	1.07	0.92-1.26	3.07E-01	1.08	0.92-1.26	2.75E-01	1.12	0.91-1.38	2.62E-01	rs2420768|rs7914572|rs11594222
H11	3	GGG	0.123	0.111	0.866	1.12	0.91-1.38	2.72E-01	1.11	0.90-1.37	3.12E-01	1.08	0.92-1.26	2.64E-01	rs2420768|rs7914572|rs11594222
H11	3	GGA	0.104	0.121	0.866	0.89	0.72-1.09	1.21E-01	0.88	0.71-1.08	9.46E-02	0.86	0.70-1.06	6.28E-02	rs2420768|rs7914572|rs11594222
H12	3	CGA	0.361	0.363	18.819	1.00	Ref.	8.97E-01	1.00	Ref.	9.90E-01	1.00	Ref.	9.12E-01	rs10458766|rs11199483|rs12570141
H12	3	AAA	0.280	0.255	18.819	1.10	0.93-1.30	8.82E-02	1.08	0.92-1.28	1.35E-01	1.09	0.92-1.29	1.43E-01	rs10458766|rs11199483|rs12570141
H12	3	AGA	0.272	0.278	18.819	0.98	0.83-1.15	6.53E-01	0.98	0.83-1.15	7.19E-01	0.98	0.83-1.16	6.62E-01	rs10458766|rs11199483|rs12570141
H12	3	AGG	0.084	0.100	18.819	0.84	0.66-1.07	9.04E-02	0.83	0.66-1.06	8.19E-02	0.87	0.68-1.10	1.55E-01	rs10458766|rs11199483|rs12570141

Identified haplotypes from genotyped variants in *PLPP4* ±100 kb upstream or downstream. A total of 45 haplotypes were identified in 12 haplotype blocks.

[a] Unadjusted logistic regression analyses

[b] Logistic regression analyses adjusted for age, sex and PC1-5

[c] Logistic regression analyses adjusted for age, sex, PC1-5 and tobacco use

Haplotypes indicating a significant association with PD at a corrected p-value threshold of 0.003 are marked in bold.

Table 4: Number of variants in *PLPP4* tested through SKAT and the results statistics at MAF threshold 1% and 5% in the genotyped data vs the imputed

	MAF < 1%		**MAF < 5%**	
	No. Variants	p-value	No. Variants	p-value
Genotyped data				
All variants	1	NA	**16**	**0.002**
Coding variants	0	NA	0	NA
Imputed data				
All variants	402	0.109	628	0.005
Coding variants	1	0.562	1	0.562

Analyses were adjusted for age at inclusion, sex and PC1-5
Results significant at Bonferroni correction are highlighted in bold

Figures

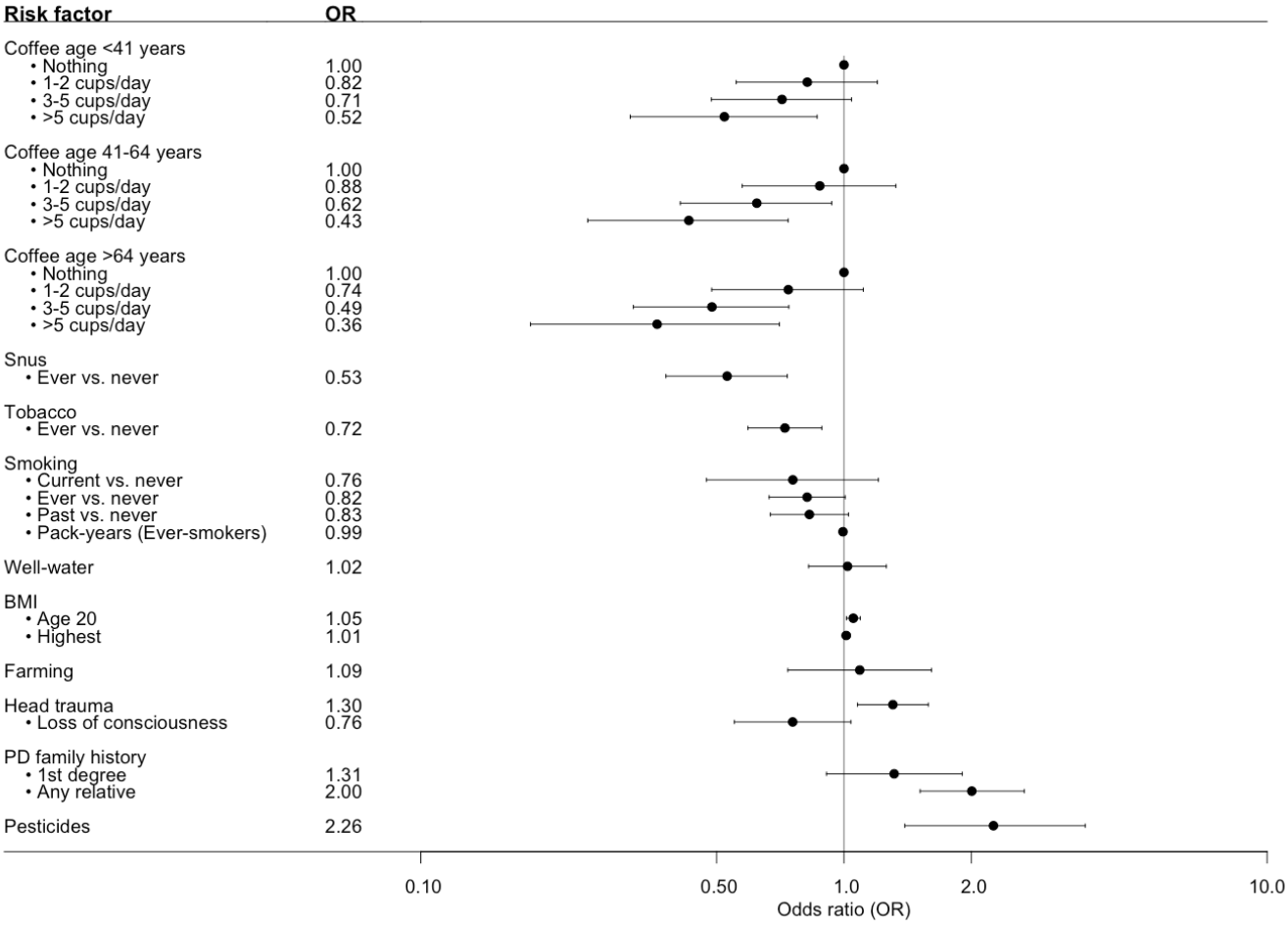

Figure 1: Forest plot over the associations between risk factors and PD in MPBC.

Showing the adjusted OR and 95% CI. Number of individuals in each analysis is found in Table S1.

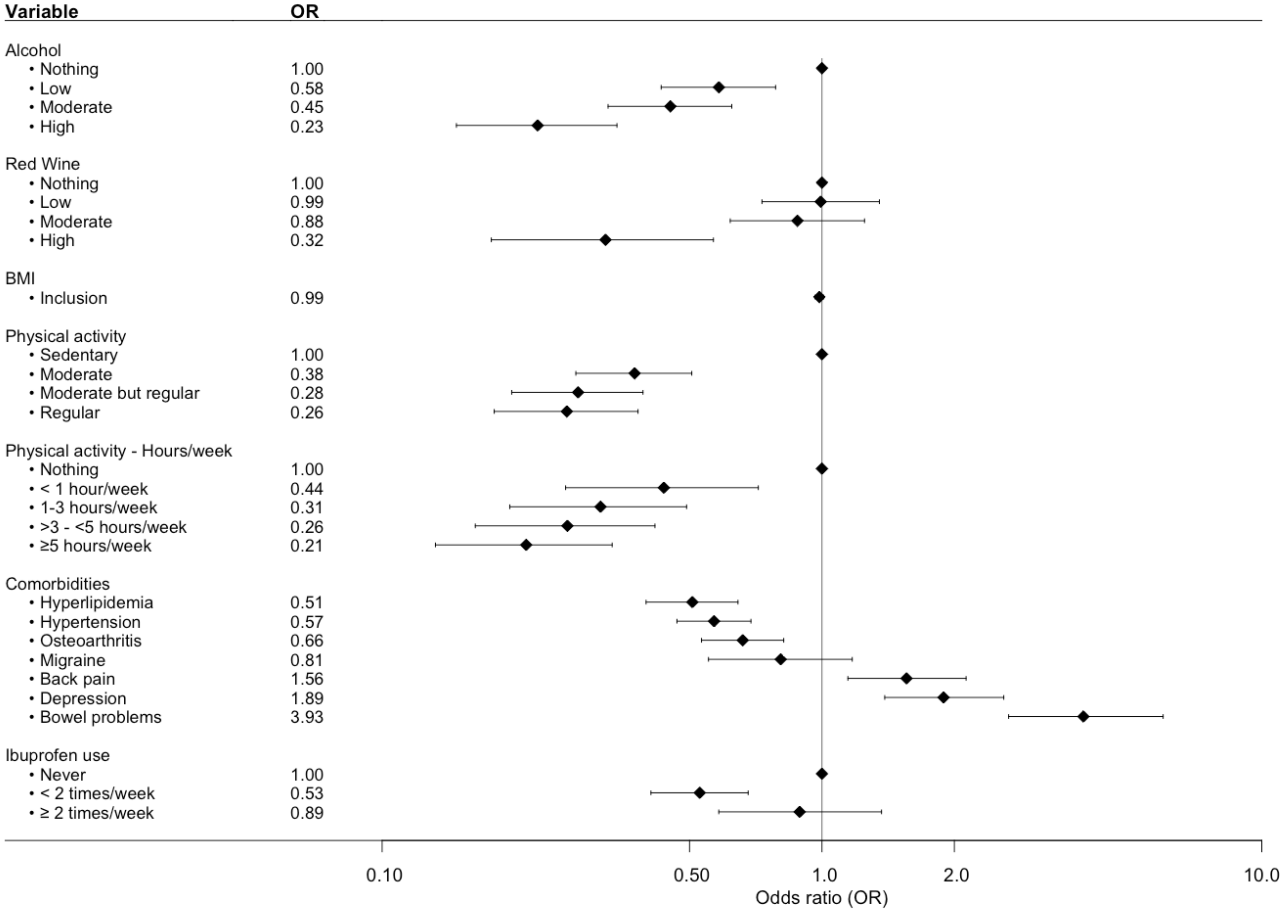

Figure 2: Forest plot over the associations between various variables and PD. Analyzed variables were obtained from questions regarding exposure/use within the past year prior to study inclusion. The plot shows OR and 95% CI adjusted by age at inclusion and sex. Number of individuals in each analysis is found in Table S2.

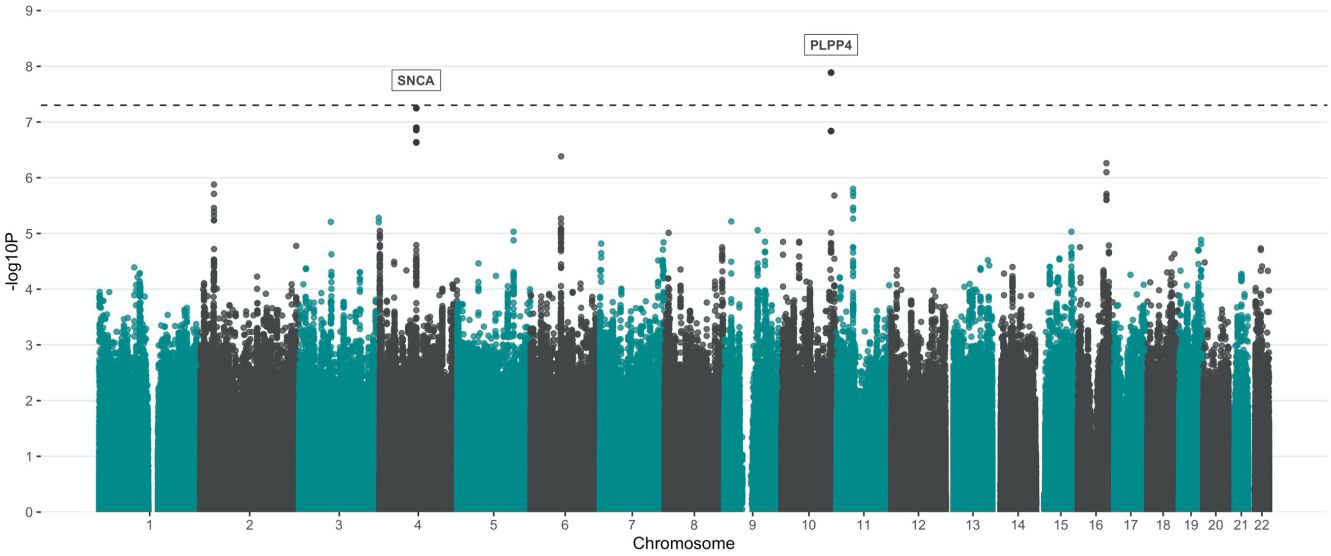

Figure 3: Manhattan plot showing the result from PD GWA analysis. A total of 5,445,841 SNPs (MAF >5%) were tested for 929 PD patients vs 935 controls. The y-axis represents the negative log (two-sided p-values) for association of variants with PD and the x-axis represent the genomic position on genome build GRCh37. The horizontal dashed line indicates the genome-wide significance level (p=5E-08).

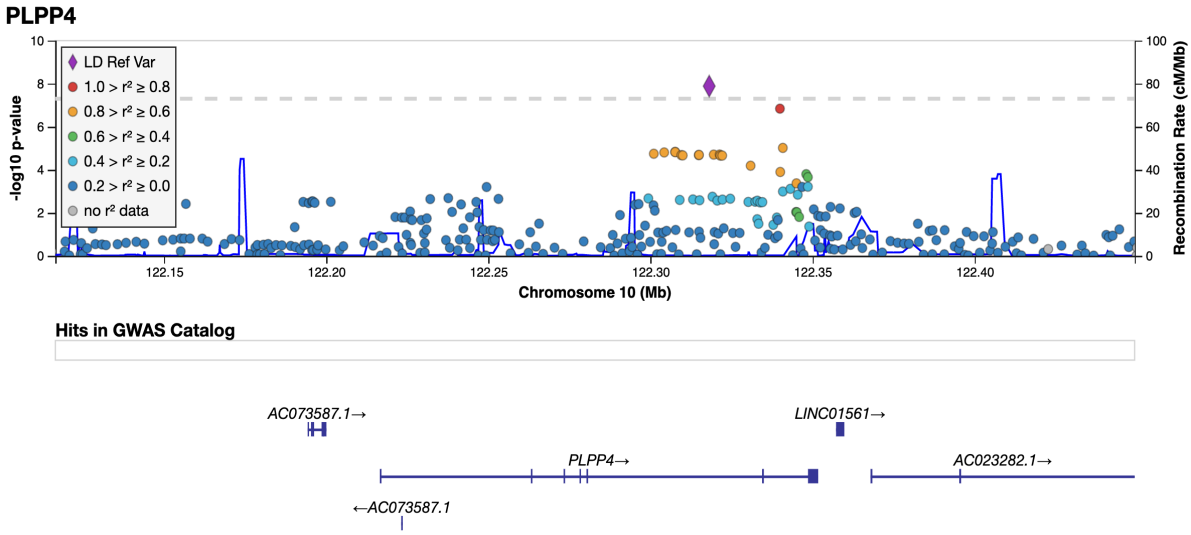

Figure 4: LocusZoom plot for PD GWAS PLPP4 loci. Imputed and genotyped variants passing QC in the *PLPP4* gene +/- 100 kb (10:122116466-122449376) mapped to genome build GRCh37. The only coding gene in the region is *PLPP4*, other include pseudogenes (AC073587.1) and long non-coding RNA (LINC01561, AC023282.1, WDR11-AS1). The variant with lowest p-value (index) is indicated as a purple diamond. Marker colors indicate the strength of LD as r^2 between the index variant and other variants in the 1000 Genomes EUR population.

1

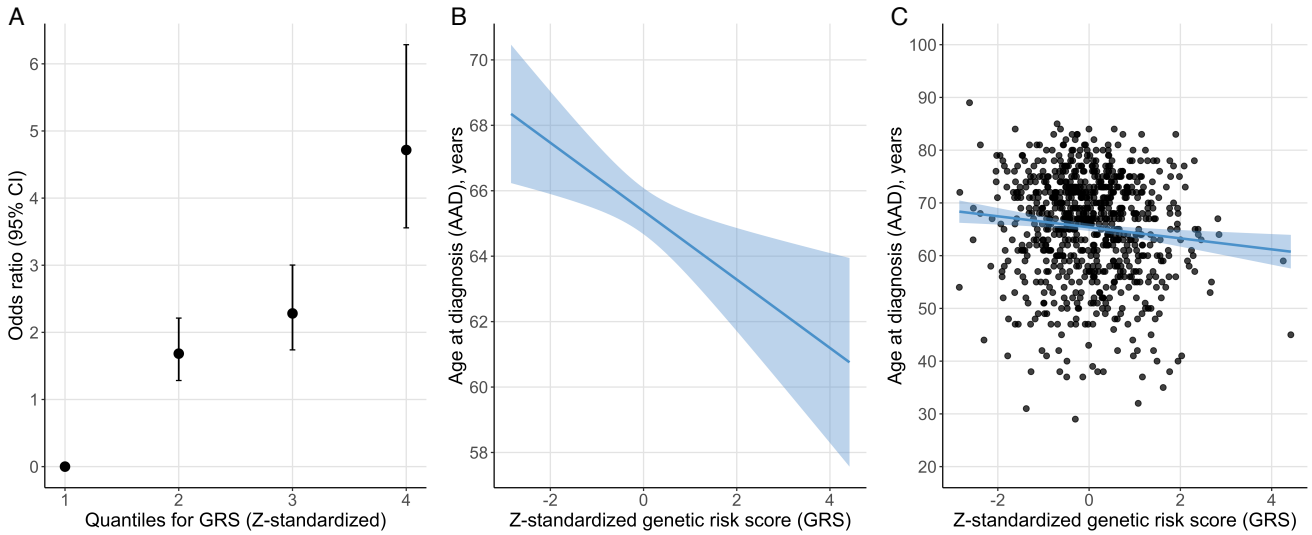

Figure 5: Genetic risk score (GRS) quartiles vs disease status (A) and age at diagnosis (AAD) (B and C). A: Odds ratio of PD status per risk quartile of the Z-standardized GRS. B: Regression line for the association between the Z-standardized GRS and AAD. The line represents the parameter estimate and the shading the 95% CI of the regression model. The model was adjusted for sex, PD family history, and PC1-5. C: AAD and Z-standardized GRS for each study participant with the regression model in plot B fitted to the plot.