Accurate long-read sequencing identified GBA variants as a major genetic risk factor in the Luxembourg Parkinson's study

Sinthuja Pachchek Peiris, MSc[1]; Zied Landoulsi, PhD[1]; Lukas Pavelka, MD[2,5] ; Claudia Schulte, MSc [3]; Elena Buena-Atienza, PhD[4]; Caspar Gross, MSc [4]; Ann-Kathrin Hauser[3]; Dheeraj Reddy Bobbili, PhD [1]; Nicolas Casadei, PhD[4]; Patrick May, PhD[1,*]; and Rejko Krüger, PhD[1,2,5,*] on behalf of the NCER-PD Consortium

[1]LCSB, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-Sur-Alzette, Luxembourg.
[2] Parkinson Research Clinic, Centre Hospitalier de Luxembourg (CHL), Luxembourg.
[3]Department of Neurodegeneration, Center of Neurology, Hertie Institute for Clinical Brain Research, German Center for Neurodegenerative Diseases, University of Tübingen, Tübingen, Germany.
[4]Institute of Medical Genetics and Applied Genomics, University of Tübingen, Tübingen, Germany; NGS Competence Center Tübingen (NCCT), University of Tübingen, Tübingen, Germany.
[5]Transversal Translational Medicine, Luxembourg Institute of Health (LIH), Strassen, Luxembourg.
*corresponding authors

ORCiD identifier:
Sinthuja Pachchek Peiris : 0000-0002-9182-4966
Zied Landoulsi : 0000-0002-2327-3904
Lukas Pavelka : 0000-0002-7721-3317
Claudia Schulte : 0000-0003-4006-1265
Elena Buena-Atienza 0000-0002-9890-1960
Caspar Gross : 0000-0002-9009-5458
Ann-Kathrin Hauser : 0009-0003-7848-5722
Dheeraj Reddy Bobbili : 0000-0002-1368-9623
Nicolas Casadei : 0000-0003-2209-0580
Patrick May : 0000-0001-8698-3770
Rejko Krueger: 0000-0002-3753-5733

Corresponding authors:
Prof. Dr. Rejko Krüger
Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg
6, avenue du Swing, L-4367 Belvaux, Luxembourg
rejko.krueger@uni.lu

Dr. Patrick May
Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg
6, avenue du Swing, L-4367 Belvaux, Luxembourg
patrick.may@uni.lu

Abstract : 187
main text : 2321
Methods : 1354
Total : 3675
figures/tables : 9
references : 55

Word count: 3675 (excluding table and references)

Running title: GBA variants in the Luxembourg Parkinson's Study

Keywords: glucocerebrosidase, Parkinson's disease, Genetics, long-read sequencing, Luxembourg

Abstract

Heterozygous variants in the glucocerebrosidase *GBA* gene are an increasingly recognized risk factor for Parkinson's disease (PD). Due to the pseudogene *GBAP1* that shares 96% sequence homology with the *GBA* coding region, accurate variant calling by array-based or short-read sequencing methods remains a major challenge in understanding the genetic landscape of *GBA*-related PD. We established a novel long-read sequencing technology for assessing the full length of the *GBA* gene. We used subsequent regression models for genotype-phenotype analyses. We sequenced 752 patients with parkinsonism and 806 healthy controls of the Luxembourg Parkinson's study. All *GBA* variants identified showed a 100% true positive rate by Sanger validation. We found 12% of unrelated PD patients carrying *GBA* variants. Three novel variants of unknown significance (VUS) were identified. Using a structure-based approach, we defined a potential risk prediction method for VUS. This study describes the full landscape of *GBA*-related parkinsonism in Luxembourg, showing a high prevalence of *GBA* variants as the major genetic risk for PD. Our approach provides an important advancement for highly accurate *GBA* variant calling, which is essential for providing access to emerging causative therapies for *GBA* carriers.

63 **1. Introduction**

64 Heterozygous variants in the glucocerebrosidase (*GBA*) gene, which encodes the enzyme β-
65 glucocerebrosidase (GCase), are increasingly recognized as the most common genetic risk
66 factor for the development of Parkinson's disease (PD). Homozygous mutations in *GBA* are
67 causative for the most frequent autosomal-recessive lysosomal storage disorder, Gaucher
68 disease (GD).[1] GD is characterized by a deficiency of the enzyme GCase which is necessary
69 to hydrolyse the β-glucosyl linkage of glucosylceramide lipide (GlcCer) in lysosomes, to yield
70 glucose and ceramide.[2]

71 The accurate variant calling in the *GBA* gene is challenging due to the presence of the highly
72 homogeneous untranslated pseudogene called *GBAP1*, which is located 16 kilobases (kbp)
73 downstream,[3] and shares 96% sequence homology within the coding region.[4] Furthermore,
74 recombination and structural chromosomal variations within and around the *GBA* locus make
75 the analysis more challenging.[5] Complex alleles, which include several point mutations, are
76 derived from a recombination between functional *GBA* and pseudogene *GBAP1*.[6] RecNciI is
77 the most prevalent recombinant allele, including the amino acid changes p.L483P and p.A495P,
78 and the synonymous variant p.V499V.[6]

79 Our study aimed at accurately assessing all coding variants in the *GBA* gene among all
80 participants of the Luxembourg Parkinson's study,[7] a case and control cohort including people
81 with PD and atypical parkinsonism. To assess the accuracy of the novel targeted *GBA*
82 sequencing method using Pacific Biosciences (PacBio)[8] technology, we compared this method
83 to genotyping with the NeuroChip array[9] and short-read whole genome sequencing (WGS) data
84 using Sanger sequencing as the gold standard for validation. We identified several types of
85 pathogenic *GBA* variants (severe, mild, and risk) and further characterized genotype-phenotype
86 associations to better understand the influence of each variant type and their effect on disease
87 severity.

88 **2. Results**

89 **2.1. Demographic and clinical characteristics**

90 A total of 752 patients (652 PD patients and 100 patients with other forms of parkinsonism)
91 and 806 HC were included (Figure 1). All participants were genotyped using NeuroChip and
92 screened for *GBA* variants using targeted PacBio method, while a subset of 72 patients was
93 screened with WGS. Among the patients, 66.4% (n = 499) were male with a mean age at disease
94 onset of 63.1 ± 16 years. The control group consisted of 52.9% (n = 426) males with a mean
95 age at assessment of 59.3 ± 12.2 years (Supplementary Table 1).

96 **2.2. Targeted PacBio sequencing showed the highest specificity to detect *GBA* variants**

97 To measure the reliability of *GBA* variant detection, we proceeded with two types of
98 comparison. We compared PacBio, WGS and NeuroChip methods for a subset of samples
99 (n=72). Then, we compared the PacBio and NeuroChip methods since they both covered most
100 of the samples (n=1558). We considered as true positives only variants validated via Sanger
101 sequencing (Supplementary Table 2).

102 First, we evaluated 72 samples screened by all three methods (Figure 2). Using the GBA-
103 targeted PacBio method, we detected six individuals carrying *GBA* variants (p.E365K (n = 3),
104 p.T408M (n = 1), p.N409S (n = 1), RecNciI , n = 1)). All the detected variants were confirmed
105 by Sanger sequencing (true positive rate (TPR) of 100%). With the WGS method, we did not
106 identify any false positive variant call, however, the WGS method failed to detect the RecNciI
107 recombinant allele in one individual (TPR of 83.3% (5/6). Using Neurochip, we detected three
108 potential *GBA* variants carriers (p.T408M (n = 1), p.N431S (n = 1), p.A215D (n=1), however,
109 only one variant (p.T408M) was subsequently confirmed by Sanger sequencing (TPR of 16.6%
110 (1/6) translating into a false detection rate (FDR) of 66.6% (2/3)). Next, we compared the
111 results from 1558 samples screened with both, the *GBA* targeted PacBio method and the
112 NeuroChip array (Figure 3). Using the GBA-targeted PacBio method, we detected 133 *GBA*
113 variants carriers, of which 100% were validated by Sanger sequencing. Through the NeuroChip
114 array, we detected 47 potential *GBA* variant carriers, among which only 36 were validated by
115 Sanger sequencing (TPR of 27% (36/133), resulting in an FDR of 23.4% (11/47)).

116 **2.3. Classification of *GBA* variants**

117 From the 1558 individuals sequenced with the GBA-targeted PacBio method, we identified
118 124 carriers with at least one *GBA* variant (Supplementary Table 3). *GBA* variants were mostly
119 heterozygous missense, one patient carried a heterozygous stop-gain variant

120 p.R398*(rs121908309), two PD patients carried a homozygous missense variant
121 p.E365K/p.E365K(rs2230288). We also detected nine different synonymous variants in exonic
122 regions (Supplementary Table 4). The variant p.T408T(rs138498426) is a splice site variant
123 (located within 2bp of the exon boundary) and classified as VUS.[10] The remaining synonymous
124 variants were not further analysed. Additionally, we identified 69 variants in intronic and UTRs
125 regions (Supplementary Table 5) with unclear pathogenic relevance, of which 35 were rare.
126 Based on Neurochip and WGS data, none of the *GBA* carriers carried pathogenic mutations in
127 other PD associated genes as defined by MDSGene.[11]
128 We classified four combinations of multiple variants per individual as severe (p.N409S-
129 p.L483P; the recombinant allele RecNciI; p.K13R-p.L483P; p.F252I-p.T408M) and one
130 combination of variants as risk (Y61H-T408M) based on the classification of the respective
131 associated pathogenic variants (Table 1).
132 Overall, we detected 12% (77/644) *GBA* variant carriers among 644 unrelated PD patients and
133 5% (34/678) in healthy control individuals. We found a frequency of 10.4% (67/644) known
134 pathogenic mutations in PD patients and 4.3% (29/678) in the control group (Table 2). Carriers
135 of severe *GBA* mutations (n=21; OR=11.4; 95% CI=[2.6, 48.8]; p=0.0010) and risk *GBA*
136 variants (n=39; OR=1.6; 95% CI=[1, 2.8]; p=0.0470) had a different risk of developing PD as
137 defined by the indicated OR.
138 The most common *GBA* variants in PD patients were the risk variants p.E365K (n=23;3.5%)
139 and p.T408M (n=17;2.6%).

140 **2.4. Genotype-phenotype associations in GBA-PD patients**

141 We characterized the clinical phenotype of severe, mild and risk *GBA* carriers and non-carriers
142 only in unrelated PD patients excluding carriers with only one synonymous or VUS variants.
143 The AAO was similar between *GBA* carriers (61.6±11.5) and non-carriers (62.5±11). Severe
144 PD_{GBA} mutations carriers showed a trend towards younger AAO compared to mild and risk
145 (severe: 58.6±13.1 vs mild: 65.4±17 vs risk: 62.5±0.3 years; p=0.29) (Table 3), with a
146 significant risk to develop early onset PD (OR=3.76;p=0.0135).
147 We compared clinical features between PD patients carrying pathogenic *GBA* variants and PD
148 patients without *GBA* variants. We found that the sense of smell was strongly impaired in
149 carriers (uncorrected p=0.0198) (Supplementary Table 6). Next, we compared patients carrying
150 variants from each category (severe, mild or risk) separately with PD patients without *GBA*
151 variants (Table 4). Carriers of severe *GBA* mutation showed more severe non-motor symptoms
152 when compared to non-GBA carriers, such as MDS-UPDRS Part I (uncorrected p=0.0088) and

153 hallucinations (uncorrected p=0.015), and also an impaired sense of smell as assessed by
154 Sniffin' Stick test (uncorrected p=0.0403). To show the deleterious impact of the severe
155 variants, we compared carriers of severe variants with patients carrying either mild or risk *GBA*
156 variants (Table 5). We observed here that severe variants carriers have more severe gait
157 disorder and depression and worse MDS-UPDRS Part I and PDQ-39. For all clinical features,
158 there were no significant associations after the correction for multiple comparisons using FDR
159 adjustment.

160 **2.5. VUS and the Glucosylceramidase structure**

161 We detected nine already reported VUS (p.K13R, p.Y61H, p.R78C, p.L213P, p.E427K,
162 p.A495P, p.H529R, p.R534C, p.T408T) and three new VUS (p.A97G, p.A215 and p.R434C).
163 According to our strategy developed for VUS *GBA* variants classification, where we assign the
164 pathogenicity based on the REVEL, the CADD, the dbscSNV scores, as well as whether the
165 patients carrying the variants. We suggest to sub-classify the variants p.Y61H, p.L213P,
166 p.A215D, and p.R434C as severe variants. The variant p.L213P changes the Leucine amino
167 acid into proline, which is known to be the 'helix breaker' amino acid that can induce a bend
168 into the protein structure[12](Supplementary Figure 1). The p.L213P and p.A215D variants are
169 in the catalytic site of the enzyme in the triose-phosphate isomerase (TIM) barrel structure. The
170 p.Y61H variant (Figure 4.A) is next in sequence and in structure to the known severe PD variant
171 p.C62W and the patient carrying this variant had an AAO of 38 years, indicating an early-onset
172 likely severe form of PD. The p. R434C variant is close to a known severe (p.V433L) and mild
173 (p.W432R, p.N435T) PD variants in the 3D structure. We propose to sub-classify the variants
174 p.H529R and p.R534C as mild, as they are both found only in PD patients. The variants
175 p.K13R, p.R78C, p.E427K, and p.A495P are sub-classified as risk variants. The variant
176 p.K13R is located in the signal peptide region. The variant p.R78C was annotated as "PD
177 susceptibility" in HGMD with deleterious impact in CADD. The variant p.E427K was
178 annotated as linked to "parkinsonism" in ClinVar and "reduced activity" in HGMD. We
179 suggest to classify the variant p.A97G as probably benign because it is localized in a coil-bend
180 structure and is not close to any known pathogenic variants.
181 The synonymous variant p.T408T was found in two cases and one healthy control individual.
182 Two established splice-site prediction scores (dbscSNV: ada_score 0.9797 and rf_score 0.85)
183 agreed in their prediction that the variant is likely to affect splicing. HGMD classified the
184 variant as disease mutation (DM) (Supplementary Table 4). Therefore, we propose to classify
185 the variant as a risk variant.

186 In total, we propose to classify four VUS variants as severe (p.Y61H, p.L213P, p.A215D, and
187 p.R434C), two as mild (p.H529R and p.R534C), five as risk (p.K13R, p.R78C, p.E427K,
188 p.A495P and p.T408T) and one as benign (p.A97G) (Figure 4.B).

189 **3. Discussion**

190 Our study showed, for the first time, the utility of targeted PacBio sequencing as a highly
191 sensitive and specific method to identify known and novel *GBA* variants. The PacBio method
192 demonstrated a very high efficiency by targeting the entire length of the *GBA* gene with 100%
193 reliability and solves the problems arising from the presence of the *GBAP1* pseudogene. The
194 effectiveness of the target PacBio method to investigate relevant genes with homologous
195 pseudogenes has also been proven in several other studies.[12–15] The comparative study that we
196 conducted with the different screening methods for *GBA* mutations will help researchers to be
197 more accurate and comprehensive implying a more critical appraisal of the results obtained by
198 NeuroChip and WGS with more false positive and false negative results.
199 *GBA* mutations were identified as the most common genetic risk factor for the development of
200 PD. A heterozygous *GBA* variant was typically observed in 4%–12% of PD patients in different
201 populations worldwide, with the highest prevalence of 20% described in Ashkenazi Jewish
202 PD patients. [16,17] Important variation is due to ethnicity, the investigated mutations and the
203 sequencing method used. Our study describes the landscape of *GBA* carriers in the studied
204 Luxembourgish population showing the high prevalence of *GBA* mutations that could be the
205 major genetic risk factor of PD in Luxembourg. The frequency of GBA mutation in PD in our
206 study was 12% and we observed a significantly higher proportion of pathogenic (severe, mild
207 and risk) *GBA* variants in PD patients compared to HC (10.4% *vs*
208 4.3%;OR=2.6;CI=[1.6,4.1],p=0.0001). Compared to previous studies, our study highlights that
209 using the new PacBio sequencing method, the Luxembourg Parkinson's study cohort showed
210 a comparable frequency of PD_{GBA} carriers reported so far in similarly sized Italian[18] and
211 Spanish[19] cohorts (Supplementary Table 7) . When comparing previous reports of *GBA* variants
212 in different populations, we want to highlight the fact that only cohorts that used full Sanger
213 sequencing were able to detect the RecNciI recombinant allele so far. This once more
214 emphasizes the accuracy of the PacBio sequencing methods for detecting rare and complex
215 *GBA* variants. Additionally, we confirmed that severe variants showed a higher OR than risk
216 variants, which supports the concept of graded risk for different *GBA* variants in PD_{GBA}
217 carriers. [20]

218 The most prevalent *GBA* variant in the Luxembourg Parkinson's study cohort was p. E365K,
219 and the frequency of this variant was similar to what was described in the Irish[20], Spanish[19] and
220 New Zealand[5] populations. It is interesting to note that homozygous carriers of the p.E326K
221 variant do not develop GD.[21] The variant is associated with PD, and multiple studies have
222 found enrichments varying from 1.60 to 3.34.[22–24]Furthermore, carriers of the risk variants
223 p.E365K and p.T408M were associated with atypical parkinsonism, as these variants were the
224 only ones also present in patients with DLB and PSP in our cohort. Whether this is simply
225 related to the higher frequency of these risk variants in the general population or does have a
226 specific impact on the phenotype needs to be determined in larger studies focusing on *GBA*
227 variants in atypical parkinsonism[25].
228 We present a concept for classifying VUS in the *GBA* gene according to the localisation in
229 relation to known variants in sequence and 3D structure, which may help to provide access to
230 future targeted therapies for these patients. Here additional *in vitro* and *ex vivo* studies are
231 needed to functionally validate the impact of these VUS on GCase function in neurons derived
232 from stem cells or in enzyme-activity assays in CSF of affected carriers of these VUS.
233 Additionally, we observed that the average AAO in PD was about four years younger in severe
234 *GBA* carriers compared to non-GBA carriers. This was also observed in previous studies, which
235 showed that PD_{GBA} patients generally have an earlier AAO compared to non-carriers with a
236 median onset in the early fifties.[26,27]
237 Recent studies have shown that PD_{GBA} carriers have a higher prevalence of cognitive
238 impairment[18,28,29] and non-motor symptoms including neuropsychiatric disturbances[18,19],
239 autonomic dysfunction[28], and sleep disturbances such as RBD[30]. Although not significant after
240 *p*-value adjustment, we found a similar trend and noticed that motor symptoms such as gait
241 disorder, non-motor such as depression and hallucinations symptoms were associated with a
242 more aggressive clinical phenotype in severe *GBA* carriers, supporting the effect of differential
243 *GBA* variant severity.[19,31]
244 In conclusion, this study showed the utility of targeted PacBio sequencing to identify known
245 and novel *GBA* variants with high accuracy. These findings offer important access to variant-
246 specific counselling. Furthermore, our study describes the full landscape of *GBA* related PD in
247 the current Luxembourgish population showing the high prevalence of *GBA* variants as the
248 major genetic risk in PD.

249 **4. Methods**

250 **4.1. Clinical Cohort**

251 At the time of analysis, the Luxembourg Parkinson's study comprised 1558 participants (752
252 patients of parkinsonism and 806 healthy controls (HC) in the frame of the National Centre for
253 Excellence in Research on Parkinson's disease program (NCER-PD).
254 All patients complied with the diagnostic criteria of typical PD or atypical parkinsonism as
255 assessed by neurological examination following the United Kingdom Parkinson's Disease
256 Society Brain Bank (UKPDSBB) diagnostic criteria[32]: 652 fulfilled the criteria for PD, 60 for
257 progressive supranuclear palsy (PSP) including corticobasal syndrome as a subtype of PSP
258 (PSP-CBS), 25 for Dementia with Lewy Body (DLB), 14 for Multiple System Atrophy (MSA),
259 and one for Fronto-temporal dementia with parkinsonism (FTDP). All patients and HC
260 underwent a comprehensive clinical assessment of motor and non-motor symptoms,
261 neuropsychological profile and medical history along with comorbidities. The clinical
262 symptoms assessed, and scales applied are defined in the Supplemental Information[33]. All
263 individuals provided written informed consent. The patients were reassessed at regular follow-
264 up visits. We considered early-onset PD patients those with age at onset (AAO) equal to or
265 younger than 45 years[34]. The genotype-phenotype analysis was based on the assessment of the
266 first visit. The study has been approved by the National Research Ethics Committee (CNER
267 Ref: 201407/13).

268 **4.2. Genetic analysis**

269 **4.2.1. NeuroChip array**

270 Genotyping was carried out on the InfiniumR NeuroChip Consortium Array[9] (v.1.0 and v1.1;
271 Illumina, San Diego, CA USA). For rare variants analysis, standard quality control (QC)
272 procedures were conducted, using PLINK v1.9[35], to remove variants if they had a low
273 genotyping rate (<95%) and Hardy-Weinberg equilibrium p-value < $1x10^{-6}$. As an additional
274 quality filter, we applied a study-wide allele frequency threshold of <1% in our cohort for rare
275 variants.

276 **4.2.2. GBA-targeted PacBio long-read amplicon sequencing**

277 The targeted *GBA* gene screening was performed by single-molecule real-time (SMRT) long
278 read sequencing[8] using Sequel II instrument (PacBio). The targeted *GBA* gene coordinates were
279 chr1:155,232,501-155,241,415 (USCS GRCh38/hg38). Long-distance PCR was performed
280 using GBA-specific primer sequences (Forward: 5'-
281 GCTCCTAAAGTTGTCACCCATACATG-3' and Reverse: 5'-
282 CCAACCTTTCTTCCTTCTTCTCAA-3')[36] and the 2x KAPA HiFi Hot Start ReadyMix

283 (Roche). For sample multiplexing, dual asymmetric barcoding was used based on a different
284 16-bp long index sequence upstream of each of the reverse and forward primers to allow the
285 generation of uniquely barcoded amplicons in one-step PCR amplification. QC was performed
286 prior to pooling. Pools of amplicons were purified with AMPure PacBio beads. A total of 1700
287 ng of purified amplicon pool was used as input for the SMRTbell library using the SMRTbell
288 Express Template Prep Kit 2.0 (PacBio). Binding of the polymerase and diffusion loading on
289 SMRTCell 8M was prepared according to SMRTLink instructions with CCS reads as
290 sequencing mode (version SMRT Link: 9.0.0.92188). We generated high-quality consensus
291 reads using the PacBio Sequel II sequencer on Circular Consensus Sequencing mode using the
292 pbccs (v6.0.0) tool. The methods replicates both strands of the target DNA.[37] We demultiplexed
293 and mapped reads from each sample to the human reference genome GRCh38 using
294 minimap2[38] from the pbmm2 package (v1.4.0)
295 (https://github.com/PacificBiosciences/pbmm2). For variant calling, we used the
296 DeepVariant[39] (1.0) with models optimized for CCS reads. Finally, we selected variants with
297 quality above 30 (QUAL>30).

298 **4.2.3. Whole genome sequencing**

299 The TruSeq Nano DNA Library Prep Kit (Illumina, San Diego, CA, USA) and MGIEasy FS
300 DNA Prep kit (BGI, China) were used according to the manufacturer's instructions to construct
301 the WGS library. Paired-end sequencing was performed with the Illumina NovaSeq 6000[40] and
302 on the MGI G400 sequencers. A QC of the raw data was performed using FastQC (version
303 0.11.9).[41] To call the variants, we used the Bio-IT Illumina Dynamic Read Analysis for
304 GENomics (DRAGEN) DNA pipeline[42] v3.8[43] with standard parameters. To select the high-
305 quality variants, we annotated and selected variants using VariantAnnotator and SelectVariants
306 modules of the Genome Analysis Toolkit (GATK 4)[44] pipeline and applied the following
307 additional filtering steps: VariantFiltration module for SNVs (QD<2, FS>60, MQ<40,
308 MQRankSum<−12, ReadPosRankSum<−8, DP<10.0, QUAL<30, VQSLOD<0, ABHet>0.75
309 or <0.25, SOR>3 and LOD<0), and insertions-deletions (QD<2, FS>200, QUAL<30,
310 ReadPosRankSum<−20, DP<10 and GQ_MEAN<20).

311 **4.3. Variant annotation and validation**

312 Variant annotation was done with ANNOVAR,[45] using the Genome Aggregation Database
313 (gnomAD r2.1)[46], the Human Gene Mutation Database (HGMD)[47] and ClinVar[48], and the
314 Combined Annotation Dependent Depletion (CADD)[49] and REVEL[50] to score the

315 pathogenicity of missense variants.[51] For variants in splice sites, we used the ada_score and
316 rf_score from dbscSNV (version 1.1)[52]. Ada_score ≥ 0.6 or rf_score ≥ 0.6 indicate that the
317 variant is likely to affect splicing.
318 Rare variants were selected according to minor allele frequency < 1% in gnomAD for the Non-
319 Finnish European (NFE) population in the 'non-neuro' gnomAD subset. Then, exonic and
320 splicing variants (+/- 2bp from the exon boundary) were selected for autosomal dominant
321 (*LRRK2*, *SNCA*, *VPS35*, *GBA*) and autosomal recessive (*PRKN*, *PINK1*, *PARK7*, *ATP13A2*)
322 PD genes. Rare variants within these genes were then confirmed by Sanger sequencing.[53]

323 **4.4. *GBA* variant nomenclature**

324 All variants in *GBA* were annotated based on GRCh37 and were numbered according to the
325 current variant nomenclature guidelines (http://varnomen.hgvs.org), based on the primary
326 translation product (NM_001005742), which includes the 39-residue signal peptide.

327 **4.5. *GBA* variant classification**

328 *GBA* variants classification was done according to the PD literature based on the work of
329 Höglinger and colleagues in 2022.[10] Exonic or splice-site variants that are not mentioned in
330 the paper were subclassified as 'severe' *GBA* variants if there were annotated as pathogenic
331 in ClinVar, otherwise they were subclassified as variant with unknown significance (VUS).
332 [55]

333 **4.6 Statistical analysis**

334 To assess the frequency of different *GBA* variant types and analyse the genotype-phenotype
335 associations in the Luxembourg Parkinson's Study, we considered only unrelated individuals
336 and retained only one proband per family. For cases, we kept the patient with the earliest AAO.
337 We excluded HC with first-degree PD relatives (parents, sibs, and offspring) and age at
338 assessment (AAA) less than 60 years, to account for age-dependent penetrance. Therewith,
339 reduce the gap of age between cases and HC. Thus, 1420 unrelated individuals (742 patients
340 and 678 HC) were selected for the statistical analysis.
341 We used regression models to assess the effect of PD_{GBA} carrier status on the clinical variables.
342 We excluded individuals carrying only VUS or synonymous variants. To this aim, we
343 performed three types of association tests: (1) all PD_{GBA} pathogenic variant carriers (severe,
344 mild and risk) *vs* $PD_{GBA\text{-non-carriers}}$, (2) for each sub-group of PD_{GBA} pathogenic variant carriers
345 *vs* $PD_{GBA\text{-non-carriers}}$, (3) severe PD_{GBA} pathogenic variant carriers *vs* combined mild and risk
346 PD_{GBA} pathogenic variant carriers. The effect of each factor was expressed as the Beta (β)

347 regression coefficient. The odds ratio (OR) along with a 95% confidence interval (CI) was used
348 to assess whether a particular exposure is a risk factor for a particular outcome. Regression
349 models were adjusted for AAA, sex, and disease duration. FDR adjusted p-value < 0.05 was
350 considered as statistically significant.

351 **4.7 Structure-based evaluation of VUS**

352 To evaluate VUS variants, we implemented a method to assign the pathogenicity based on the
353 REVEL and CADD scores for missense variants and the dbscSNV scores (ada_score and
354 rf_score) for splice variants according to dbNFSP definition[54], as well as whether the patients
355 carrying the variants.We reclassified a VUS (1) as 'severe' if the variant was present only in
356 patients and with deleterious effect in all scores or present only in patients with early onset PD,
357 (2) as 'mild' if the variant was present only in patients and with tolerated effect in all scores,
358 (3) as ' risk' if present in patients and HCs or with tolerated and deleterious effect in either
359 score or annotated as 'PD susceptibility' in HGMD, and (4) as 'benign' if present only in HC.
360 We mapped the known pathogenic missense variants and newly identified VUS identified in
361 our cohort together with all reported population variants from gnomAD onto the *GBA* protein
362 sequence and the 3D structure. We used an X-ray structure of GCase at 2.0 Å resolution (PDB
363 structure accession code 1ogs; https://www.rcsb.org/) (Supplementary Figure 2). The analysis
364 of the 3D structure was carried out by PyMOL (http://www.pymol.org). VUS were evaluated
365 as risk variant if they were 2bp positions away in sequence or had a C-alpha distance of less
366 than 5 ångström in 3D from another known pathogenic variant similar to the approach used in
367 Johannesen et al.[55]

368 **5. Data availability**

369 The dataset for this manuscript is not publicly available as it is linked to the Luxembourg
370 Parkinson's Study and its internal regulations. Any requests for accessing the dataset can be
371 directed to request.ncer-pd@uni.lu.

372 **6. References**

373 1. Hruska KS, LaMarca ME, Scott CR, Sidransky E. Gaucher disease: mutation
374 and polymorphism spectrum in the glucocerebrosidase gene (GBA). *Hum*
375 *Mutat*. 2008;29(5):567-583. doi:10.1002/HUMU.20676
376 2. Arturo-Terranova D, Giraldo LJM, Satizábal JM. Frequency of gba gene
377 variants in complex disease patients in Southwestern Colombia. *Genetics and*
378 *Molecular Research*. 2021;20(2). doi:10.4238/gmr18818
379 3. Horowitz M, Wilder S, Horowitz Z, Reiner O, Gelbart T, Beutler E. The human
380 glucocerebrosidase gene and pseudogene: structure and evolution. *Genomics*.
381 1989;4(1):87-96. doi:10.1016/0888-7543(89)90319-4
382 4. Do J, McKinney C, Sharma P, Sidransky E. Glucocerebrosidase and its
383 relevance to Parkinson disease. *Mol Neurodegener*. 2019;14(1).
384 doi:10.1186/s13024-019-0336-2
385 5. Graham OEE, Pitcher TL, Liau Y, et al. Nanopore sequencing of the
386 glucocerebrosidase (GBA) gene in a New Zealand Parkinson's disease cohort.
387 *Parkinsonism Relat Disord*. 2020;70:36-41.
388 doi:10.1016/j.parkreldis.2019.11.022
389 6. Zimran A, Horowitz M. *RecTL: A Complex Allele of the Glucocerebrosidase*
390 *Gene Associated With a Mild Clinical Course of Gaucher Disease*.; 1994.
391 7. Hipp G, Vaillant M, Diederich NJ, et al. The Luxembourg Parkinson's Study: A
392 Comprehensive Approach for Stratification and Early Diagnosis. *Front Aging*
393 *Neurosci*. 2018;10. doi:10.3389/FNAGI.2018.00326
394 8. Korlach J, Bjornson KP, Chaudhuri BP, et al. Real-time DNA sequencing from
395 single polymerase molecules. *Methods Enzymol*. 2010;472:431-455.
396 doi:10.1016/S0076-6879(10)72001-2
397 9. Blauwendraat C, Faghri F, Pihlstrom L, et al. NeuroChip, an updated version of
398 the NeuroX genotyping platform to rapidly screen for variants associated with
399 neurological diseases. *Neurobiol Aging*. 2017;57:247.e9-247.e13.
400 doi:10.1016/J.NEUROBIOLAGING.2017.05.009
401 10. Höglinger G, Schulte C, Jost WH, et al. GBA-associated PD: chances and
402 obstacles for targeted treatment strategies. *J Neural Transm*. Published online
403 May 31, 2022. doi:10.1007/s00702-022-02511-7
404 11. Lill CM, Mashychev A, Hartmann C, et al. Launching the movement disorders
405 society genetic mutation database (MDSGene). *Mov Disord*. 2016;31(5):607-
406 609. doi:10.1002/MDS.26651
407 12. Qiao W, Yang Y, Sebra R, et al. Long-Read Single Molecule Real-Time Full
408 Gene Sequencing of Cytochrome P450-2D6. *Hum Mutat*. 2016;37(3):315-323.
409 doi:10.1002/HUMU.22936
410 13. Buermans HPJ, Vossen RHAM, Anvar SY, et al. Flexible and Scalable Full-
411 Length CYP2D6 Long Amplicon PacBio Sequencing. *Hum Mutat*.
412 2017;38(3):310-316. doi:10.1002/HUMU.23166
413 14. Borràs DM, Vossen RHAM, Liem M, et al. Detecting PKD1 variants in
414 polycystic kidney disease patients by single-molecule long-read sequencing.
415 *Hum Mutat*. 2017;38(7):870-879. doi:10.1002/HUMU.23223
416 15. Frans G, Meert W, van der Werff Ten Bosch J, et al. Conventional and Single-
417 Molecule Targeted Sequencing Method for Specific Variant Detection in
418 IKBKG while Bypassing the IKBKGP1 Pseudogene. *J Mol Diagn*.
419 2018;20(2):195-202. doi:10.1016/J.JMOLDX.2017.10.005

16. Ruskey JA, Greenbaum L, Roncière L, et al. Increased yield of full GBA sequencing in Ashkenazi Jews with Parkinson's disease. *Eur J Med Genet*. 2019;62(1). doi:10.1016/j.ejmg.2018.05.005
17. Gan-Or Z, Amshalom I, Kilarski LL, et al. *Differential Effects of Severe vs Mild GBA Mutations on Parkinson Disease*.; 2015.
18. Petrucci S, Ginevrino M, Trezzi I, et al. GBA-Related Parkinson's Disease: Dissection of Genotype–Phenotype Correlates in a Large Italian Cohort. *Movement Disorders*. 2020;35(11). doi:10.1002/mds.28195
19. Jesús S, Huertas I, Bernal-Bernal I, et al. GBA variants influence motor and non-motor features of Parkinson's disease. *PLoS One*. 2016;11(12). doi:10.1371/journal.pone.0167749
20. Olszewska DA, McCarthy A, Soto-Beasley AI, et al. Association Between Glucocerebrosidase Mutations and Parkinson's Disease in Ireland. *Front Neurol*. 2020;11. doi:10.3389/fneur.2020.00527
21. Duran R, Mencacci NE, Angeli A V., et al. The glucocerobrosidase E326K variant predisposes to Parkinson's disease, but does not cause Gaucher's disease. *Movement Disorders*. 2013;28(2):232-236. doi:10.1002/mds.25248
22. Ran C, Brodin L, Forsgren L, et al. Strong association between glucocerebrosidase mutations and Parkinson's disease in Sweden. *Neurobiol Aging*. 2016;45. doi:10.1016/j.neurobiolaging.2016.04.022
23. Davis MY, Johnson CO, Leverenz JB, et al. Association of GBA mutations and the E326K polymorphism with motor and cognitive progression in parkinson disease. *JAMA Neurol*. 2016;73(10):1217-1224. doi:10.1001/jamaneurol.2016.2245
24. Berge-Seidl V, Pihlstrøm L, Maple-Grødem J, et al. The GBA variant E326K is associated with Parkinson's disease and explains a genome-wide association signal. *Neurosci Lett*. 2017;658. doi:10.1016/j.neulet.2017.08.040
25. Picillo M, Petrucci S, Valente EM, et al. Progressive Supranuclear Palsy–Like Phenotype in a GBA E326K Mutation Carrier. *Mov Disord Clin Pract*. 2017;4(3):444-446. doi:10.1002/mdc3.12406
26. Blauwendraat C, Heilbron K, Vallerga CL, et al. Parkinson's disease age at onset genome-wide association study: Defining heritability, genetic loci, and α-synuclein mechanisms. *Mov Disord*. 2019;34(6):866-875. doi:10.1002/MDS.27659
27. Sidransky E, Nalls MA, Aasly JO, et al. Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. *N Engl J Med*. 2009;361(17):1651-1661. doi:10.1056/NEJMOA0901281
28. Brockmann K, Srulijes K, Hauser AK, et al. *GBA-Associated PD Presents with Nonmotor Characteristics*.; 2011.
29. Setó-Salvia N, Pagonabarraga J, Houlden H, et al. Glucocerebrosidase mutations confer a greater risk of dementia during Parkinson's disease course. *Movement Disorders*. 2012;27(3). doi:10.1002/mds.24045
30. Krohn L, Ruskey JA, Rudakou U, et al. GBA variants in REM sleep behavior disorder: a multicenter study. doi:10.1101/19010991
31. Brockmann K, Quadalti C, Lerche S, et al. Association between CSF alpha-synuclein seeding activity and genetic status in Parkinson's disease and dementia with Lewy bodies. *Acta Neuropathol Commun*. 2021;9(1). doi:10.1186/S40478-021-01276-6

32. Litvan I, Bhatia KP, Burn DJ, et al. SIC task force appraisal of clinical diagnostic criteria for parkinsonian disorders. *Movement Disorders*. 2003;18(5). doi:10.1002/mds.10459
33. Pavelka L, Rauschenberger A, Landoulsi Z, et al. ARTICLE Age at onset as stratifier in idiopathic Parkinson's disease-effect of ageing and polygenic risk score on clinical phenotypes. doi:10.1038/s41531-022-00342-7
34. Gustavsson EK, Trinh J, Mckenzie M, et al. Genetic Identification in Early Onset Parkinsonism among Norwegian Patients. Published online 2017. doi:10.1002/mdc3.12501
35. Purcell S, Neale B, Todd-Brown K, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559-575. doi:10.1086/519795
36. Leija-Salazar M, Sedlazeck FJ, Toffoli M, et al. Evaluation of the detection of GBA missense mutations and other variants using the Oxford Nanopore MinION. *Mol Genet Genomic Med*. 2019;7(3). doi:10.1002/mgg3.564
37. Rhoads A, Au KF. PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics*. 2015;13(5). doi:10.1016/j.gpb.2015.08.002
38. Li H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34(18):3094-3100. doi:10.1093/bioinformatics/bty191
39. Poplin R, Chang PC, Alexander D, et al. A universal snp and small-indel variant caller using deep neural networks. *Nat Biotechnol*. 2018;36(10):983. doi:10.1038/nbt.4235
40. Illumina. NovaSeq 6000 Sequencing System. *770-2016-025-H*. 2016;4(February).
41. Andrews S. FastQC. *Babraham Bioinformatics*. Published online 2010.
42. Miller NA, Farrow EG, Gibson M, et al. A 26-hour system of highly sensitive whole genome sequencing for emergency management of genetic diseases. *Genome Med*. 2015;7(1). doi:10.1186/s13073-015-0221-8
43. Illumina. Illumina DRAGEN Bio-IT Platorm. *User Guide*. 2019;(February).
44. Mark AD, Eric B, Ryan P, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Published online 2011. doi:10.1038/ng.806
45. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16). doi:10.1093/NAR/GKQ603
46. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans , Genome Aggregation Database Consortium. *434 | Nature* |. 2020;581:19. doi:10.1038/s41586-020-2308-7
47. Stenson PD, Ball E v, Mort M, et al. Human Gene Mutation Database (HGMD s): 2003 Update. *Hum Mutat*. 2003;21:577-581. doi:10.1002/humu.10212
48. Landrum MJ, Lee JM, Riley GR, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. doi:10.1093/nar/gkt1113
49. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res*. 2019;47. doi:10.1093/nar/gky1016
50. Tian Y, Pesaran T, Chamberlin A, et al. REVEL and BayesDel outperform other in silico meta-predictors for clinical variant classification. *Sci Rep*. 2019;9(1). doi:10.1038/s41598-019-49224-8

516 51. Ioannidis NM, Rothstein JH, Pejaver V, et al. REVEL: An Ensemble Method
517 for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet*.
518 2016;99(4):877-885. doi:10.1016/J.AJHG.2016.08.016
519 52. Jian X, Boerwinkle E, Liu X. In silico prediction of splice-altering single
520 nucleotide variants in the human genome. *Nucleic Acids Res*.
521 2014;42(22):13534-13544. doi:10.1093/nar/gku1206
522 53. Sanger F, Nicklen S, Coulson AR. *DNA Sequencing with Chain-Terminating*
523 *Inhibitors (DNA Polymerase/Nucleotide Sequences/Bacteriophage 4X174)*. Vol
524 74.; 1977. https://www.pnas.org
525 54. Liu X, Li C, Mou C, Dong Y, Tu Y. dbNSFP v4: a comprehensive database of
526 transcript-specific functional predictions and annotations for human
527 nonsynonymous and splice-site SNVs. *Genome Med*. 2020;12(1):1-8.
528 doi:10.1186/S13073-020-00803-9/FIGURES/4
529 55. Johannesen KM, Liu Y, Koko M, et al. Genotype-phenotype correlations in
530 SCN8A-related disorders reveal prognostic and therapeutic implications. *Brain*.
531 2022;145(9):2991-3009. doi:10.1093/BRAIN/AWAB321
532
533

7. Acknowledgments

We would like to thank all participants of the Luxembourg Parkinson's Study for their important support of our research. Furthermore, we acknowledge the joint effort of the National Centre of Excellence in Research on Parkinson's Disease (NCER-PD) Consortium members from the partner institutions Luxembourg Centre for Systems Biomedicine, Luxembourg Institute of Health, Centre Hospitalier de Luxembourg, and Laboratoire National de Santé generally contributing to the Luxembourg Parkinson's Study as listed below. The work presented here was funded by the Luxembourg National Research (FNR/NCER13/BM/11264123), the PEARL program (FNR/P13/6682797 to RK), MotaSYN (12719684 to RK), MAMaSyn (to RK), MiRisk-PD (C17/BM/11676395 to RK, PM), the FNR/DFG Core INTER (ProtectMove, FNR11250962 to PM), and the PARK-QC DTU (PRIDE17/12244779/PARK-QC to RK, SP).

ON BEHALF OF THE NCER-PD CONSORTIUM

Geeta ACHARYA [2], Gloria AGUAYO [2], Myriam ALEXANDRE [2], Muhammad ALI [1], Wim AMMERLANN [2], Giuseppe ARENA [1], Rudi BALLING [1], Michele BASSIS [1], Katy BEAUMONT [2], Regina BECKER [1], Camille BELLORA [2], Guy BERCHEM [3], Daniela BERG [11], Alexandre BISDORFF 5, Ibrahim BOUSSAAD [1], Kathrin BROCKMANN [11], Jessica CALMES [2], Lorieza CASTILLO [2], Gessica CONTESOTTO [2], Nico DIEDERICH [3], Rene DONDELINGER [5], Daniela ESTEVES [2], Guy FAGHERAZZI [2], Jean-Yves FERRAND [2], Manon GANTENBEIN [2], Thomas GASSER [11], Piotr GAWRON [1], Soumyabrata GHOSH [1], Marijus GIRAITIS [2,3], Enrico GLAAB [1], Elisa GÓMEZ DE LOPE [1], Jérôme GRAAS [2], Mariella GRAZIANO [17], Valentin GROUES [1], Anne GRÜNEWALD [1], Wei GU [1], Gaël HAMMOT [2], Anne-Marie HANFF [2], Linda HANSEN [1],3, Maxime HANSEN [1,3], Michael HENEKA [1], Estelle HENRY [2], Sylvia HERBRINK [6], Sascha HERZINGER [1], Michael HEYMANN [2], Michele HU [8], Alexander HUNDT [2], Nadine JACOBY [18], Jacek JAROSLAW LEBIODA [1], Yohan JAROZ [1], Quentin KLOPFENSTEIN [1], Jochen KLUCKEN [1,2],3, Rejko KRÜGER [1,2,3], Pauline LAMBERT [2], Zied LANDOULSI [1], Roseline LENTZ [7], Inga LIEPELT [11], Robert LISZKA [14], Laura LONGHINO [3], Victoria LORENTZ [2], Paula Cristina LUPU [2], Clare MACKAY [10], Walter MAETZLER [15], Katrin MARCUS [13], Guilherme MARQUES [2], Tainá M MARQUES [1], Patricia MARTINS CONDE [1], Patrick MAY [1], Deborah MCINTYRE [2], Chouaib MEDIOUNI [2], Francoise MEISCH [1], Myriam MENSTER [2], Maura MINELLI [2], Michel MITTELBRONN [1,4], Brit MOLLENHAUER [12], Friedrich MÜHLSCHLEGEL [4], Romain NATI [3], Ulf NEHRBASS [2], Sarah NICKELS [1], Beatrice NICOLAI [3], Jean-Paul NICOLAY [19], Marek OSTASZEWSKI [1], Clarissa P. da C. GOMES [1], Sinthuja PACHCHEK [1], Claire PAULY [1,3], Laure PAULY [1], Lukas PAVELKA [1,3], Magali PERQUIN [2], Rosalina RAMOS LIMA [2], Armin RAUSCHENBERGER [1], Rajesh RAWAL [1], Dheeraj REDDY BOBBILI [1], Kirsten ROOMP [1], Eduardo ROSALES [2], Isabel ROSETY [1], Estelle SANDT [2], Stefano SAPIENZA [1], Venkata SATAGOPAM [1], Margaux SCHMITT [2], Sabine SCHMITZ [1], Reinhard SCHNEIDER [1], Jens SCHWAMBORN [1], Jean-Edouard SCHWEITZER [1], Amir SHARIFY [2], Ekaterina SOBOLEVA [1], Kate SOKOLOWSKA [2], Olivier TERWINDT [1],3, Hermann THIEN [2], Elodie THIRY [3], Rebecca TING JIIN LOO [1], Joana TORRE [2], Christophe TREFOIS [1], Johanna TROUET [2], Olena TSURKALENKO [2], Michel VAILLANT [2], Mesele VALENTI [2], Carlos VEGA [1], Liliana VILAS BOAS [3], Maharshi VYAS [1], Richard WADE-MARTINS [1], Paul WILMES [1], Evi WOLLSCHEID-LENGELING [1], Gelani ZELIMKHANOV [3]

[1]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg
[2]Luxembourg Institute of Health, Strassen, Luxembourg
3 Centre Hospitalier de Luxembourg, Strassen, Luxembourg
4 Laboratoire National de Santé, Dudelange, Luxembourg

5 Centre Hospitalier Emile Mayrisch, Esch-sur-Alzette, Luxembourg
6 Centre Hospitalier du Nord, Ettelbrück, Luxembourg
7 Parkinson Luxembourg Association, Leudelange, Luxembourg
8 Oxford Parkinson's Disease Centre, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK
9 Oxford Parkinson's Disease Centre, Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, UK
10 Oxford Centre for Human Brain Activity, Wellcome Centre for Integrative Neuroimaging, Department of Psychiatry, University of Oxford, Oxford, UK
11 Center of Neurology and Hertie Institute for Clinical Brain Research, Department of Neurodegenerative Diseases, University Hospital Tübingen, Tübingen, Germany
12 Paracelsus-Elena-Klinik, Kassel, Germany
13 Ruhr-University of Bochum, Bochum, Germany
14 Westpfalz-Klinikum GmbH, Kaiserslautern, Germany
15 Department of Neurology, University Medical Center Schleswig-Holstein, Kiel, Germany
16 Department of Neurology Philipps, University Marburg, Marburg, Germany
17 Association of Physiotherapists in Parkinson's Disease Europe, Esch-sur-Alzette, Luxembourg
18 Private practice, Ettelbruck, Luxembourg
19 Private practice, Luxembourg, Luxembourg

8. Author contribution

1. Research project: A. Conception, B. Organization, C. Execution;
2. Statistical Analysis: A. Design, B. Execution, C. Review and Critique;
3. Manuscript Preparation: A. Writing of the first draft, B. Review and Critique;
4. Genetic data: A. Sequencing Execution, B. Analysis;
5. Data collection: A. Participation, B. Exportation, C. Curation

SP: 1A, 1B, 1C, 2A, 2B, 2C, 3A, 3B, 4B,5B, 5C
ZL: 1C, 2A , 2B, 2C,3A, 3B, 5C
LP: 1C, 2C, 3B, 5A, 5B, 5C
CS: 2C, 3B, 4A
EBA: 2C, 3B, 4A
CG: 2C, 3B, 4A
AKH: 2C, 3B, 4A
DRB: 2C, 3B
NC: 3B, 4A
PM: 1A, 1B, 1C, 2A, 2C, 3B, 4B, 5C
RK: 1A, 1B, 1C, 2A, 2C, 3B, 5A, 5C

9. Financial Disclosures and Conflict of Interest

The authors declare that there are no conflicts of interest relevant to this work.

10. Funding Sources

The National Centre of Excellence in Research on Parkinson's Disease (NCER-PD) is funded by the Luxembourg National Research Fund (FNR/NCER13/BM/11264123), the PEARL program (FNR/P13/6682797 to RK), MotaSYN (12719684 to RK), MAMaSyn (to RK), MiRisk-PD (C17/BM/11676395 to RK, PM), the FNR/DFG Core INTER (ProtectMove, FNR11250962 to PM), and the PARK-QC DTU (PRIDE17/12244779/PARK-QC to RK, SP).

622 **11. Tables**

623 **Table 1. Distribution of *GBA* variants in the Luxembourg Parkinson's study.**

Subclassification	nucleotide - protein changes	Subjects	PD n=652	Parkinsonism Patients n=100	Healthy controls n=806
Severe	c.115+1G>A	2	1		1
	p.G234W	1	1		
	p.G241R	2	2		
	p.H294Q	1			1
	p.R398*	1	1		
	p.G416S	1	1		
	p.L483P	6	5		1
	p.R502H	1	1		
	p.N409S; p.L483P	1	1		
	p.L483P; p.A495P; p.V499V (*RecNciI*)	5	4		1
	p.K13R; p.L483P	1	1		
	p.F252I; p.T408M	1	1		
Mild	p.N409S	10	7		3
Risk	p.E365K	42	21+2[a]	1 DLB + 2 PSP	16
	p.T408M	28	15	1 DLB	12
	p.Y61H; p.T408M	1	1		
VUS	p.K13R	4	2		2
	p.Y61H	1	1		
	p.R78C	2			2
	p.A97G (new VUS)	1			1
	p.P161S	2	2		
	p.L213P	1	1		
	p.A215D (new VUS)	1	1		
	p.E427K	2	1		1
	p.R434C (new VUS)	1	1		
	p.H529R	1	1		
	p.R534C	1	1		
	p.A495P; p.V499V	3	1		2
	p.T408T	3	2		1

624 All variants were identified in the heterozygous state except in two individuals for p.E365K ([a]
625 Homozygous state). Abbreviations: GBA, glucocerebrosidase gene; PD, Parkinson's Disease and
626 Parkinson's Disease with Dementia; PSP, Progressive Supranuclear Palsy; DLB, Dementia with Lewy
627 Body; VUS, Variants of unknown significance.

628 **Table 2. Frequency of GBA mutations in the Luxembourg Parkinson's study.**

Diagnosis	sub-classification of *GBA* mutations	Subjects	All GBA-Carrier (n)	Pathogenic GBA-Carrier	OR (95%CI)	*p*-values
PD		644	**77 (12%)**	**67 (10.4%)**	2.6 (1.6 to 4.1)	0.0001*
	severe			21 (3.2%)	11.4 (2.6 to 48.8)	0.0010*
	mild			7 (1.1%)	3.5 (0.7 to 17.2)	0.1137
	risk			39 (6%)	1.6 (1 to 2.8)	0.0470*
PSP	risk	59		2 (3.4%)	0.9 (0.2 to 3.9)	0.8941
DLB	risk	24		2 (8.3%)	2.2 (0.5 to 10)	0.2908
Healthy controls		678	34 (5%)	29 (4.3%)	-	-
	severe			2	-	-
	mild			2	-	-
	risk			25	-	-

629 After excluding the first-degree family members interrelated in the cohort and the healthy controls with
630 young age of assessment (< 60 AAA) with first degree PD relatives. GBA, glucocerebrosidase gene;
631 PD, Parkinson's Disease and Parkinson's Disease with Dementia; PSP, Progressive Supranuclear Palsy;
632 DLB, Dementia with Lewy Body. ORs are given with the 95% CI; Statistically significant results
633 highlighted in bold and red with * (p-value < 0.05).

634 **Table 3. Demographic data for the PD patients in the Luxembourgish cohort separated by GBA**
635 **mutation status.**

Features	All pathogenic variants (n = 67)	Severe (n = 21)	Mild (n = 7)	Risk (n = 39)	Noncarriers (n = 561)
AAA, mean (SD)	66.5 (±10.2) [OR=0.37; p=0.4713]	65.1 (±10.2) [OR=0.1; p=0.3315]	67.1 (±15.6) [OR=0.68; p=0.9275]	67.1 (±9.2) [OR=0.67; p=0.8231]	67.5 (±10.9)
Sex, Male % (n)	40 (59.7%) [OR=0.71; p=0.1882]	13 (61.9%) [OR=0.77; p=0.5762]	5 (71.4%) [OR=1.19; p=0.8356]	22 (56.4%) [OR=0.62; p=0.149]	380 (67.7%)
AAO, mean (SD)	61.6 (±11.5) [OR=0.42; p=0.5668]	58.6 (±13.1) [OR=0.02; p=0.1383]	65.4 (±17.0) [OR=19.33; p=0.511]	62.5 (±9.3) [OR=1.07; p=0.9704]	62.5 (±11.8)
AAO ≤ 45	8 (11.9%) [OR=1.63; p=0.2301]	**5 (23.8%) [OR=3.76; p=0.0135*]**	2 (28.6%) [OR=4.82; p=0.0648]	1 (2.6%) [OR=0.32; p=0.2626]	43 (7.7%)
Disease Duration, mean (SD)	4.7 (±4.8) [OR=0.7; p=0.7074]	6.4 (±4.7) [OR=3.99; p=0.2338]	1.7 (±1.4) [OR=0.04; p=0.0979]	4.4 (±4.9) [OR=0.56; p=0.4966]	5 (±5.2)
Family History, N (%)	25 (37.3%) [OR=1.5; p=0.1137]	8 (38.1%) [OR=1.58; p=0.3167]	2 (28.6%) [OR=1.03; p=0.9726]	15 (38.5%) [OR=1.61; p=0.1651]	157 (28%)

636 After excluding the first-degree family members interrelated in the cohort, the healthy controls with
637 young age of assessment (< 60 AAA) with first degree PD relatives and synonymous and VUS
638 variants carriers. Data are given as mean (SD) or N (%). Significance level for comparison is $p < 0.05$.
639 AAA, age at assessment in years ; AAO, Age at onset in years.
640

Table 4. Clinical characteristics of PD classified by GBA mutation status.

		PD *GBA* carrier															PD GBA non-carrier
		SEVERE					MILD					RISK					
Type of data	Clinical characteristics and scales	PD_{GBA} (n=21)	missing values (%)	β (95%)	*p*-value	adj *p*-value	PD_{GBA} (n=7)	missing values (%)	β (95%)	*p*-value	adj *p*-value	PD_{GBA} (n=39)	missing values (%)	β (95%)	*p*-value	adj *p*-value	N=570
Motor symptoms/scales	H&Y, mean (SD)	2.4 (±0.8)	3 (0.5%)	0.18 (-0.11 to 0.48)	0.2254	0.5635	2.0 (±0.6)	3 (0.5%)	-0.0 (-0.51 to 0.51)	0.9963	0.9992	2.2 (±0.8)	3 (0.5%)	0.07 (-0.16 to 0.29)	0.5503	0.9914	2.2 (±0.8)
	MDS-UPDRS II, mean (SD)	12.6 (±4.4)	13 (2.2%)	0.44 (-2.79 to 3.67)	0.7892	0.9441	10.1 (±6.4)	13 (2.3%)	0.86 (-4.72 to 6.44)	0.7629	0.9992	11.1 (±8.6)	13 (2.2%)	0.12 (-2.31 to 2.55)	0.9223	0.9914	11.4 (±8.3)
	MDS-UPDRS III, mean (SD)	34.8 (±15.7)	14 (2.4%)	-0.25 (-7.18 to 6.69)	0.9441	0.9441	26.2 (±9.7)	13 (2.3%)	-4.81 (-16.98 to 7.36)	0.4387	0.9992	33.3 (±17.7)	12 (2.0%)	-0.49 (-5.42 to 4.45)	0.8469	0.9914	34.6 (±16.2)
	MDS-UPDRS IV, mean (SD)	3.0 (±4.5)	7 (1.2%)	0.85 (-0.49 to 2.18)	0.214	0.5635	0.1 (±0.4)	7 (1.2%)	-0.64 (-2.88 to 1.6)	0.5737	0.9992	1.2 (±2.3)	7 (1.2%)	-0.43 (-1.4 to 0.53)	0.3779	0.8221	1.7 (±3.4)
	Dyskinesias, n (%)	5 (23.8%)	0	0.62 (-0.54 to 1.78)	0.2937	0.6425	0	-	-	-	-	4 (10.3%)	0	-0.01 (-1.15 to 1.13)	0.9878	0.9914	67 (11.9%)
	Falls, n (%)	5 (23.8%)	0	0.37 (-0.7 to 1.45)	0.4932	0.7193	0	-	-	-	-	7 (17.9%)	0	0.12 (-0.78 to 1.02)	0.7938	0.9914	98 (17.5%)
	Gait Disorder, n (%)	16 (76.2%)	0	0.9 (-0.13 to 1.93)	0.0873	0.5441	3 (42.9%)	0	-0.26 (-1.78 to 1.26)	0.7364	0.9992	18 (46.2%)	0	-0.32 (-1.0 to 0.35)	0.3503	0.8221	314 (56.0%)
	FOG, n (%)	8 (38.1%)	0	0.65 (-0.34 to 1.65)	0.1966	0.5635	0	-	-	-	-	7 (17.9%)	0	-0.2 (-1.14 to 0.73)	0.673	0.9914	126 (22.5%)
	Restless leg syndrome, n (%)	2 (9.5%)	0	0.06 (-1.44 to 1.56)	0.9413	0.9441	2 (28.6%)	0	1.62 (-0.07 to 3.31)	0.0605	0.8843	6 (15.4%)	0	0.71 (-0.22 to 1.64)	0.1324	0.8221	46 (8.2%)
	Motor fluctuation, n (%)	5 (23.8%)	0	0.12 (-1.03 to 1.27)	0.8427	0.9441	0	-	-	-	-	5 (12.8%)	0	-0.22 (-1.25 to 0.82)	0.6795	0.9914	95 (16.9%)
Non-motor symptoms/scales	BDI, mean (SD)	12.4 (±5.7)	28 (4.8%)	2.02 (-0.99 to 5.03)	0.1879	0.5635	8.0 (±3.5)	29 (5.1%)	-1.16 (-6.73 to 4.41)	0.6826	0.9992	8.0 (±5.7)	29 (4.8%)	-1.9 (-4.16 to 0.36)	0.0994	0.8221	9.9 (±7.1)
	MDS-UPDRS Part I, mean (SD)	**15.0 (±6.5)**	**15 (2.6%)**	**3.92 (0.99 to 6.86)**	**0.0088***	**0.2625**	8.6 (±2.4)	15 (2.6%)	-0.81 (-5.81 to 4.19)	0.7506	0.9992	9.5 (±6.8)	15 (2.5%)	-0.94 (-3.12 to 1.23)	0.3946	0.8221	**10.6 (±7.0)**
	PDSS, mean (SD)	98.3 (±20.9)	44 (7.6%)	-4.42 (-15.25 to 6.41)	0.424	0.6745	110.9 (±13.4)	43 (7.6%)	2.6 (-15.45 to 20.65)	0.778	0.9992	104.0 (±24.9)	45 (7.5%)	-1.3 (-9.38 to 6.78)	0.7527	0.9914	104.7 (±24.9)
	SCOPA-AUT, mean (SD)	17.1 (±8.0)	33 (5.7%)	1.67 (-1.76 to 5.1)	0.3393	0.6598	13.1 (±7.1)	32 (5.6%)	-0.26 (-5.98 to 5.46)	0.9284	0.9992	14.1 (±7.8)	33 (5.5%)	-0.67 (-3.2 to 1.86)	0.6038	0.9914	15.0 (±8.1)
	Sniffin's stick test, mean (SD)	**6.4 (±3.6)**	**7 (1.2%)**	**-1.56 (-3.05 to -0.07)**	**0.0403***	**0.4702**	6.6 (±2.1)	7 (1.2%)	-1.58 (-4.12 to 0.95)	0.2212	0.9992	7.4 (±4.0)	8 (1.3%)	-0.66 (-1.79 to 0.48)	0.2558	0.8221	**7.8 (±3.6)**
	SAS, mean (SD)	15.8 (±5.2)	34 (5.8%)	2.0 (-0.5 to 4.5)	0.1173	0.5441	12.6 (±3.8)	33 (5.8%)	-1.48 (-5.66 to 2.71)	0.4892	0.9992	13.1 (±6.3)	35 (5.8%)	-0.77 (-2.66 to 1.11)	0.4228	0.8221	14.0 (±5.7)
	MoCA, mean (SD)	24.0 (±4.7)	12 (2.1%)	-0.75 (-2.55 to 1.05)	0.4134	0.6745	25.6 (±3.7)	12 (2.1%)	0.89 (-2.15 to 3.92)	0.5664	0.9992	24.9 (±4.0)	14 (2.3%)	0.16 (-1.2 to 1.52)	0.8165	0.9914	24.4 (±4.5)
	Constipation, n (%)	10 (47.6%)	0	0.04 (-0.85 to 0.92)	0.9338	0.9441	5 (71.4%)	0	1.38 (-0.28 to 3.03)	0.1029	0.9004	14 (35.9%)	0	-0.33 (-1.02 to 0.36)	0.3518	0.8221	251 (44.7%)
	Dysphagia, n (%)	4 (19.0%)	0	-0.46 (-1.57 to 0.65)	0.4182	0.6745	1 (14.3%)	0	-0.53 (-2.66 to 1.6)	0.6273	0.9992	10 (25.6%)	0	0.0 (-0.75 to 0.76)	0.9914	0.9914	146 (26.0%)
	Insomnia, n (%)	6 (28.6%)	0	-0.06 (-1.04 to 0.92)	0.9072	0.9441	4 (57.1%)	0	1.42 (-0.11 to 2.95)	0.068	0.8843	7 (17.9%)	0	-0.56 (-1.41 to 0.28)	0.1898	0.8221	154 (27.5%)
	Orthostatism, n (%)	5 (23.8%)	0	-0.36 (-1.39 to 0.67)	0.4915	0.7193	4 (57.1%)	0	1.37 (-0.14 to 2.89)	0.0758	0.8843	15 (38.5%)	0	0.43 (-0.25 to 1.11)	0.2136	0.8221	166 (29.6%)
	Urinary incontinence, n (%)	6 (28.6%)	0	-0.1 (-1.08 to 0.88)	0.8443	0.9441	2 (28.6%)	0	0.14 (-1.54 to 1.82)	0.8729	0.9992	17 (43.6%)	0	0.63 (-0.04 to 1.31)	0.0666	0.8221	171 (30.5%)
	Hallucinations, n (%)	**8 (38.1%)**	**0**	**1.16 (0.23 to 2.09)**	**0.015***	**0.2625**	2 (28.6%)	0	1.18 (-0.49 to 2.85)	0.1664	0.9707	6 (15.4%)	0	0.05 (-0.87 to 0.97)	0.9088	0.9914	**87 (15.5%)**
	Excessive daytime sleepiness, n (%)	9 (42.9%)	0	0.45 (-0.45 to 1.35)	0.3302	0.6598	0	-	-	-	-	14 (35.9%)	0	0.3 (-0.4 to 0.99)	0.4041	0.8221	176 (31.4%)
	ICD, n (%)	2 (9.5%)	0	-0.26 (-1.79 to 1.27)	0.7402	0.9441	0	-	-	-	-	4 (10.3%)	0	0.21 (-0.89 to 1.31)	0.7051	0.9914	55 (9.8%)
	Syncope, n (%)	2 (9.5%)	0	0.84 (-0.7 to 2.38)	0.2842	0.6425	1 (14.3%)	0	1.68 (-0.54 to 3.91)	0.1377	0.9639	3 (7.7%)	0	0.55 (-0.71 to 1.82)	0.3891	0.8221	27 (4.8%)
	RBDSQ, mean (SD)	10 (47.6%)	42 (7.2%)	0.61 (-0.33 to 1.55)	0.2021	0.5635	1 (14.3%)	42 (7.4%)	-0.53 (-2.7 to 1.64)	0.6317	0.9992	14 (35.9%)	43 (7.2%)	0.38 (-0.35 to 1.12)	0.3052	0.8221	171 (30.5%)
Other clinical outcomes	LEDD (mg/day), mean (SD)	690.5 (±457.9)	19 (3.3%)	118.74 (-34.03 to 271.52)	0.1277	0.5441	324.7 (±224.7)	20 (3.5%)	-68.27 (-348.63 to 212.09)	0.6332	0.9992	496.6 (±443.2)	20 (3.3%)	8.54 (-106.24 to 123.32)	0.884	0.9914	514.7 (±404.7)
	PDQ-39, mean (SD)	52.0 (±26.3)	49 (8.4%)	9.17 (-1.69 to 20.04)	0.098	0.5441	26.1 (±17.7)	49 (8.6%)	-7.06 (-25.57 to 11.44)	0.4545	0.9992	34.9 (±27.2)	51 (8.5%)	-4.58 (-12.85 to 3.69)	0.278	0.8221	39.6 (±26.8)
	DBS, n (%)	3 (14.3%)	0	1.44 (-0.13 to 3.01)	0.0729	0.5441	0	0	-17.62 (-34180.02 to 34144.78)	0.9992	0.9992	1 (2.6%)	0	-0.06 (-2.23 to 2.12)	0.9588	0.9914	24 (4.3%)
Comorbidities	Diabetes, n (%)	2 (9.5%)	0	0.19 (-1.33 to 1.71)	0.8093	0.9441	1 (14.3%)	0	0.14 (-2.07 to 2.36)	0.8992	0.9992	5 (12.8%)	0	0.47 (-0.54 to 1.47)	0.3607	0.8221	55 (9.8%)
	Hypercholesterolemia, n (%)	6 (28.6%)	0	-0.43 (-1.41 to 0.55)	0.386	0.6745	2 (28.6%)	0	-0.66 (-2.35 to 1.04)	0.4466	0.9992	17 (43.6%)	0	0.15 (-0.52 to 0.81)	0.6631	0.9914	228 (40.6%)
	Cardiovascular disease, n (%)	1 (4.8%)	0	-1.55 (-3.6 to 0.51)	0.1399	0.5441	1 (14.3%)	0	-0.63 (-2.86 to 1.59)	0.577	0.9992	8 (20.5%)	0	0.12 (-0.72 to 0.95)	0.7863	0.9914	118 (21.0%)
	Arterial hypertension, n (%)	9 (42.9%)	0	0.11 (-0.81 to 1.03)	0.8192	0.9441	2 (28.6%)	0	-0.89 (-2.62 to 0.84)	0.314	0.9992	12 (30.8%)	0	-0.57 (-1.28 to 0.15)	0.1206	0.8221	250 (44.6%)
	Traumatic Brain Injury, n (%)	5 (23.8%)	0	0.15 (-0.88 to 1.18)	0.7798	0.9441	0	-	-	-	-	6 (15.4%)	0	-0.44 (-1.34 to 0.45)	0.3343	0.8221	124 (22.1%)

Comparison of each type (severe, mild, risk) of *GBA* mutations and its association with clinical characterization. We used regression models (linear and logistic). Data are given as mean and standard deviation (SD) for continuous clinical outcomes and as percentage for binary clinical outcomes. Models

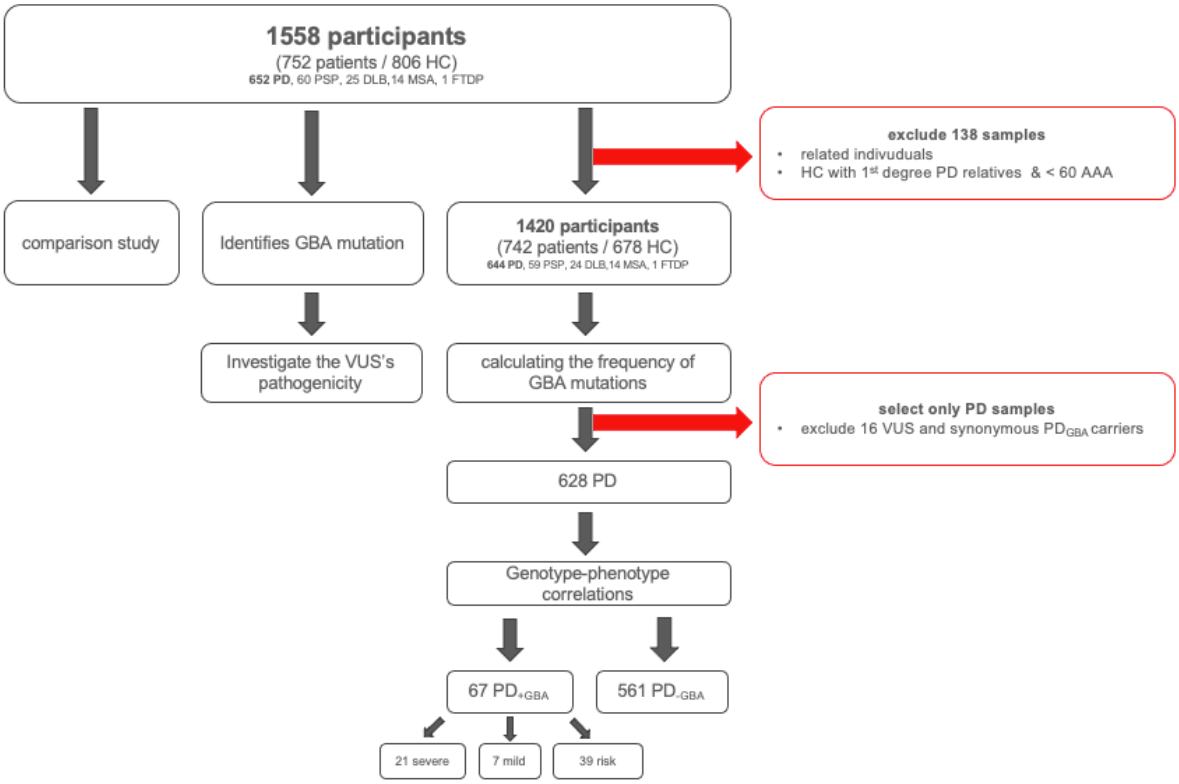
644 adjusted for sex, age at assessment, and disease duration. Beta (β) regression coefficient are given with the 95% CI. Statistically significant results highlighted
645 in bold with (*) sign and red (p-value < 0.05). Abbreviation : p-value, unadjusted p-value; adj p-value, corrected for multiple comparisons using FDR
646 adjustment; AAO, age at onset; H&Y, Hoehn & Yahr; MDS-UPDRS, Movement Disorders Society - Unified Parkinson's Disease Rating Scale; FOG,
647 freezing of gait; BDI, Beck Depression Inventory; PDSS, Panic Disorder Severity Scale; SCOPA-AUT, Scales for Outcomes in Parkinson's Disease-
648 Autonomic questionnaire; SAS, Starkstein apathy scale; MoCA, Montreal Cognitive Assessment; ICD, impulse control disorder; RBDSQ, REM Sleep
649 Behavior Disorder Screening Questionnaire; LEDD, L-dopa equivalent daily dose (mg/day); PDQ-39, Parkinson's Disease quality of life Questionnaire;
650 DBS, Presence of treatment by Deep Brain Stimulation

Table 5. The deleterious impact of severe GBA-PD carriers in comparison with mild and risk and their clinical characteristics.

Type of data	Clinical characteristics and scales	PD_{GBA} carrier		missing values (%)	β (95%)	*p*-value	adj *p*-value
		severe (n=21)	mild + risk (N=46)				
Motor symptoms/scales	H&Y, mean (SD)	2.4 (±0.8)	2.2 (±0.8)	0	0.14 (-0.24 to 0.52)	0.4719	0.7549
	MDS-UPDRS II, mean (SD)	12.6 (±4.4)	10.9 (±8.2)	0	0.24 (-3.18 to 3.65)	0.8922	0.9706
	MDS-UPDRS III, mean (SD)	34.8 (±15.7)	32.4 (±16.9)	3 (4.5%)	0.65 (-7.69 to 8.98)	0.8793	0.9706
	MDS-UPDRS IV, mean (SD)	3.0 (±4.5)	1.0 (±2.2)	0	1.33 (-0.17 to 2.82)	0.082	0.574
	Dyskinesias, n (%)	5 (23.8%)	4 (8.7%)	0	0.65 (-1.09 to 2.39)	0.4621	0.7549
	Falls, n (%)	5 (23.8%)	7 (15.2%)	0	0.24 (-1.29 to 1.77)	0.7594	0.9601
	Gait Disorder, n (%)	**16 (76.2%)**	**21 (45.7%)**	**0**	**1.49 (0.25 to 2.73)**	**0.0188***	**0.2193**
	FOG, n (%)	8 (38.1%)	7 (15.2%)	0	0.79 (-0.73 to 2.32)	0.3091	0.7549
	Restless leg syndrome, n (%)	2 (9.5%)	8 (17.4%)	0	-0.98 (-2.81 to 0.85)	0.2952	0.7549
	Motor fluctuation, n (%)	5 (23.8%)	5 (10.9%)	0	0.28 (-1.36 to 1.92)	0.7348	0.9601
Non-motor symptoms/scales	**BDI, mean (SD)**	**12.4 (±5.7)**	**8.0 (±5.4)**	**2 (3.0%)**	**4.03 (1.08 to 6.98)**	**0.0074***	**0.1295**
	MDS-UPDRS Part I, mean (SD)	**15.0 (±6.5)**	**9.3 (±6.3)**	**0**	**4.91 (1.8 to 8.02)**	**0.0019***	**0.0665**
	PDSS, mean (SD)	98.3 (±20.9)	105.1 (±23.5)	3 (4.5%)	-2.79 (-14.9 to 9.33)	0.6521	0.9129
	SCOPA-AUT, mean (SD)	17.1 (±8.0)	13.9 (±7.6)	2 (3.0%)	2.61 (-1.62 to 6.85)	0.2269	0.7542
	Sniffin's stick test, mean (SD)	6.4 (±3.6)	7.3 (±3.7)	1 (1.5%)	-0.86 (-2.75 to 1.02)	0.3695	0.7549
	SAS, mean (SD)	15.8 (±5.2)	13.0 (±6.0)	3 (4.5%)	2.02 (-0.94 to 4.97)	0.1817	0.7542
	MoCA, mean (SD)	24.0 (±4.7)	25.0 (±3.9)	2 (3.0%)	-0.24 (-2.4 to 1.92)	0.8292	0.9706
	Constipation, n (%)	10 (47.6%)	19 (41.3%)	0	0.12 (-1.02 to 1.26)	0.8394	0.9706
	Dysphagia, n (%)	4 (19.0%)	11 (23.9%)	0	-0.44 (-1.83 to 0.95)	0.5348	0.8138
	Insomnia, n (%)	6 (28.6%)	11 (23.9%)	0	-0.02 (-1.27 to 1.23)	0.9767	0.9836
	Orthostatism, n (%)	5 (23.8%)	19 (41.3%)	0	-0.61 (-1.83 to 0.61)	0.327	0.7549
	Urinary incontinence, n (%)	6 (28.6%)	19 (41.3%)	0	-0.76 (-1.97 to 0.44)	0.2156	0.7542
	Hallucinations, n (%)	8 (38.1%)	8 (17.4%)	0	1.03 (-0.24 to 2.3)	0.1127	0.6485
	Excessive daytime sleepiness, n (%)	9 (42.9%)	14 (30.4%)	0	0.41 (-0.71 to 1.52)	0.4745	0.7549
	ICD, n (%)	2 (9.5%)	4 (8.7%)	0	-0.02 (-1.89 to 1.85)	0.9836	0.9836
	Syncope, n (%)	2 (9.5%)	4 (8.7%)	0	0.1 (-1.72 to 1.92)	0.9151	0.9706
	RBDSQ, mean (SD)	10 (47.6%)	15 (32.6%)	4 (6.0%)	0.48 (-0.68 to 1.64)	0.4197	0.7549
Other clinical outcomes	LEDD (mg/day), mean (SD)	690.5 (±457.9)	473.1 (±422.4)	2 (3.0%)	66.4 (-107.09 to 239.89)	0.4531	0.7549
	PDQ-39, mean (SD)	**52.0 (±26.3)**	**33.5 (±25.9)**	**2 (3.0%)**	**12.77 (0.45 to 25.09)**	**0.0422***	**0.3692**
	DBS, n (%)	3 (14.3%)	1 (2.2%)	0	0.76 (-2.17 to 3.69)	0.6122	0.8928
Comorbidities	Diabetes, n (%)	2 (9.5%)	6 (13.0%)	0	-0.26 (-2.0 to 1.48)	0.7681	0.9601
	Hypercholesterolemia, n (%)	6 (28.6%)	19 (41.3%)	0	-0.47 (-1.63 to 0.69)	0.4269	0.7549
	Cardiovascular disease, n (%)	1 (4.8%)	9 (19.6%)	0	-1.71 (-3.92 to 0.5)	0.1297	0.6485
	Arterial hypertension, n (%)	9 (42.9%)	14 (30.4%)	0	0.66 (-0.49 to 1.81)	0.2586	0.7542
	Traumatic Brain Injury, n (%)	5 (23.8%)	6 (13.0%)	0	0.82 (-0.55 to 2.2)	0.2385	0.7542

We used regression models (linear and logistic). Data are given as mean and standard deviation (SD) for continuous clinical outcomes and as percentage for binary clinical outcomes. Models adjusted for sex, age at assessment, and disease duration. Beta (β) regression coefficient are given with the 95% CI. Statistically significant results highlighted in bold with (*) sign and red (p-value < 0.05). Abbreviation : p-value, unadjusted p-value; adj p-value, corrected for multiple comparisons using FDR adjustment; AAO, age at onset; H&Y, Hoehn & Yahr; MDS-UPDRS, Movement Disorders Society - Unified Parkinson's Disease Rating Scale; FOG, freezing of gait; BDI, Beck Depression Inventory; PDSS, Panic Disorder Severity Scale; SCOPA-AUT, Scales for Outcomes in Parkinson's Disease-Autonomic questionnaire; SAS, Starkstein apathy scale; MoCA, Montreal Cognitive Assessment; ICD, impulse control disorder; RBDSQ, REM Sleep Behavior Disorder Screening Questionnaire; LEDD, L-dopa equivalent daily dose (mg/day); PDQ-39, Parkinson's Disease quality of life Questionnaire; DBS, Presence of treatment by Deep Brain Stimulation.

666 **12. Figures**



667

668 **Figure 1: Description of the study dataset and methodology.**

A.

dbSNP	protein change	Sample count	PacBio	WGS	NeuroChip
rs2230288	p.E365K	3	3 TP	3 TP	
rs75548401	p.T408M	1	TP	TP	TP
rs76763715	p.N409S	1	TP	TP	
rs77738682	p.N431S	1			FP
	p.A215D	1			FP
	*RecNcil**	1	TP		

B.

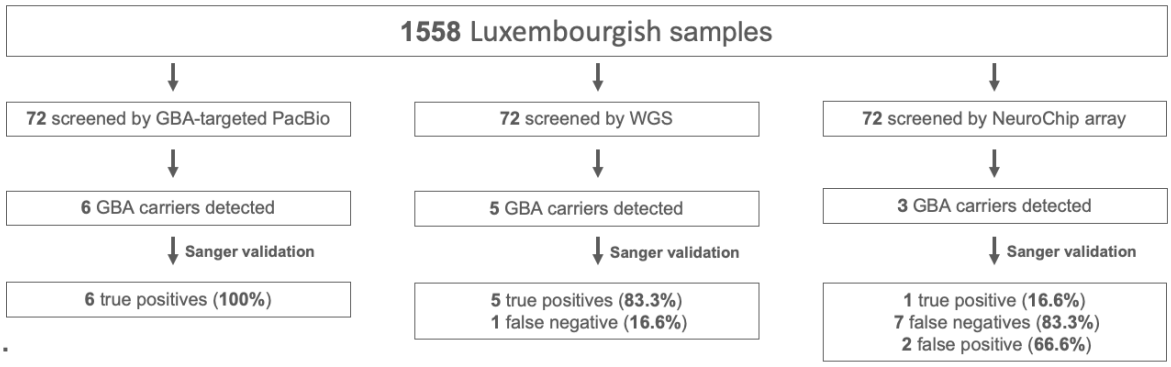
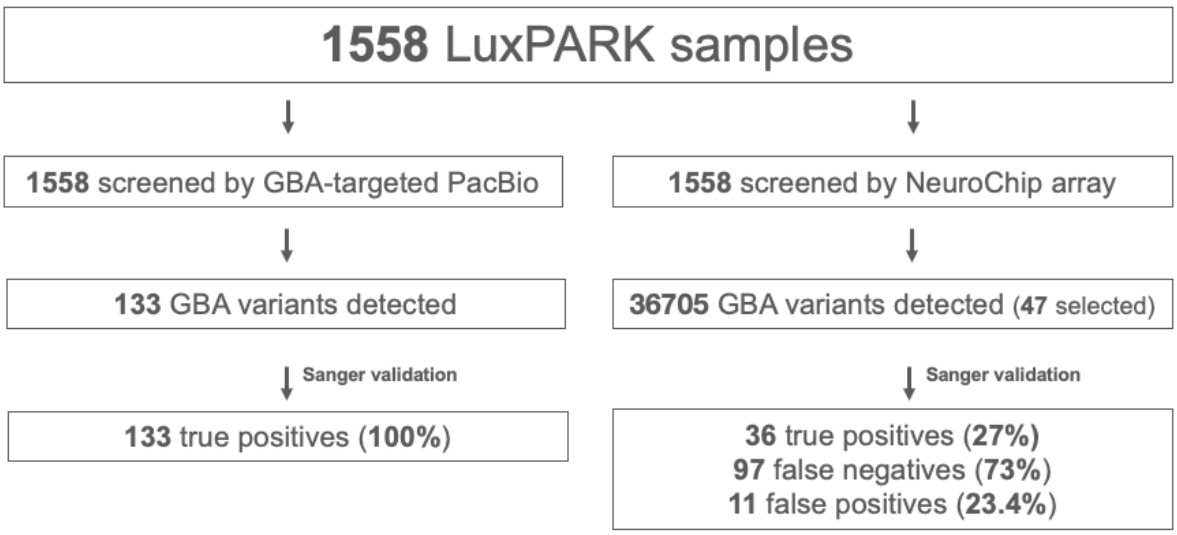


Figure 2. Comparison of variant calls from PacBio, WGS and NeuroChip genotyping data using 72 matched samples for the *GBA* gene and validated by Sanger sequencing.

A)**Rec*Ncil *(p.L483P; p.A495P; p.V499V)*; Sanger sequencing results : TP, true positive ; FP, false positive. Sample count gives total number of samples carrying the variant found by each method.

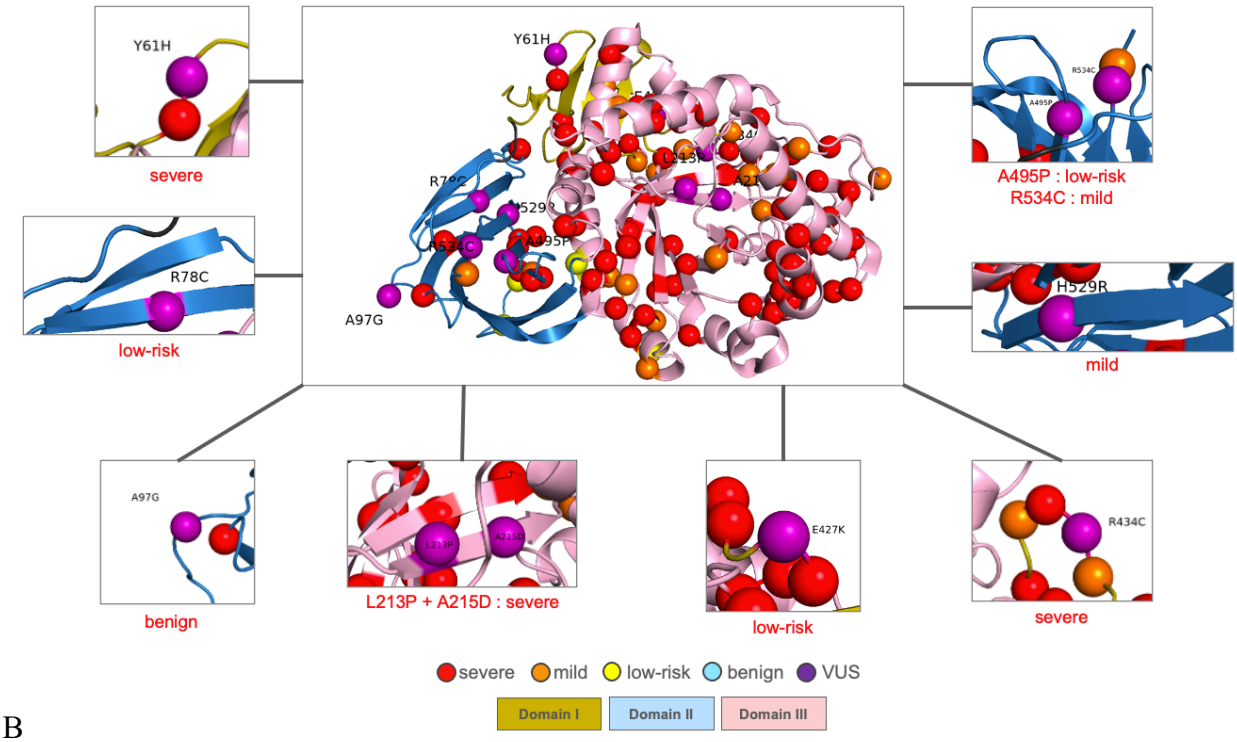
B) Comparative study of *GBA* variants detection by the GBA-targeted PacBio and NeuroChip array methods in the Luxembourg Parkinson's study. Due to overrepresented variants with the NeuroChip array, we applied for the detected variants a study-wide threshold of 1% in our cohort.



677

678 **Figure 3**: **Comparative study of GBA variants detection by the GBA-targeted PacBio and**
679 **NeuroChip array methods in the Luxembourg Parkinson's study.** Due to overrepresented
680 variants with the NeuroChip array, we applied for the detected variants a study-wide threshold
681 of 1% in our cohort.

A



B

Prediction	Subjects		protein change	nucleotide change	dbSNP	Höglinger et al.	ClinVar Significance	ClinVar	HGMD	HGMD interpretation	REVEL	CADD	GnomAD NFE	3D	domain
	PD (AA0)	HC (AAA)													
risk	2 (50/66)	3 (37/38/66)	K13R	c.A38G	rs150466109	VUS	Benign	GD	DM	GD	T	T	ultra-rare	/	/
severe	1 (38)		Y61H	c.T181C	rs1266341749	-	-	-	-	-	T	T	ultra-rare	coil-loop	I
risk		2 (56/58)	R78C	c.C232T	rs146774384	-	-	-	DM	PD susceptibility	T	D	ultra-rare	β-sheet	II
benign		1 (60)	A97G	c.C290G	-	-	-	-	-	-	T	T	-	coil-bend	II
severe	1 (56)		L213P	c.T638C	-	VUS	-	-	DM	PD	D	D	-	β-sheet	III
severe	1 (68)		A215D	c.C644A	-	-	-	-	DM	GD	D	D	-	β-sheet	III
risk	1 (65)	1 (70)	E427K	c.G1279A	rs149171124	VUS	Uncertain significance	Parkinsonism	DM	Reduced activity	T	T	ultra-rare	coil-turn	I
severe	1 (62)		R434C	c.C1300T	rs747284798	-	-	-	DM	GD 1	D	D	-	coil-loop	I
risk	1 (60)	2 (38/61)	A495P	c.G1483C	rs368060	-	Benign	GD	DM	-	T	T	ultra-rare	β-sheet	II
mild	1 (77)		H529R	c.A1586G	-	VUS	-	-	DM	PD	T	T	-	β-sheet	II
mild	1 (78)		R534C	c.C1600T	rs146519305	-	-	-	-	-	T	T	ultra-rare	coil-loop	II
risk	2 (61/64)	1 (61)	T408T	c.G1224A	rs138498426	VUS	Uncertain significance	GD	DM	PD	-	-	ultra-rare	buried residue	III

Figure 4. Sub-classification of VUS found in the Luxembourg Parkinson's study.
A) *GBA* missense and stop gain variants mapped onto the three-dimensional structure of GCase. Domain 1 is shown in dark yellow, domain 2 in blue, and domain 3 in pink. Variants classified as severe are coloured red, mild are coloured orange, risk in yellow and VUS are coloured purple. B) GBA, glucocerebrosidase gene; GD, Gaucher's disease; PD, Parkinson's disease. HGMD, The Human Gene Mutation Database; REVEL, Rare Exome Variant Ensemble Learner; CADD, Combined Annotation Dependent Depletion; gnomAD, The Genome Aggregation Database. DM, Disease causing mutation; D, Deleterious; T, Tolerate. Variants classified as severe are coloured red, mild are coloured orange, risk in yellow and VUS are coloured purple.