# Nanopore sequencing of the glucocerebrosidase (*GBA*) gene in a New Zealand Parkinson's disease cohort

**Authors:**

Graham, O.E.E. [1], Pitcher, T.L. [2, 3], Liau, Y. [1], Miller, A.L. [1], Dalrymple-Alford, J.C. [3, 4], Anderson, T.J. [2, 3], Kennedy, M.A[1].

**Author affiliations:**

[1.]Gene Structure and Function Laboratory, Department of Pathology and Biomedical Science, University of Otago, Christchurch, New Zealand. [2.] Department of Medicine, University of Otago, Christchurch, Christchurch, New Zealand. [3.]New Zealand Brain Research Institute, Christchurch, New Zealand. [4] Department of Psychology, University of Canterbury, Christchurch, New Zealand

**Keywords:**

Movement disorder, Gaucher disease, genetics, mutation detection, neurogenetics, GridION

**Highlights:**

- Validated nanopore sequencing of the *GBA* gene for mutation detection.
- Found 10.4% (21/229) of a New Zealand Parkinson's disease cohort carry *GBA* risk mutations.
- E365K and T408M are the most abundant mutations in the cohort.
- Three novel mutations were detected including a putative splicing variant.
- Single molecule reads of *GBA* amplicons revealed 74 different *GBA* haplotypes.

## Abstract

Mutations in the gene for glucocerebrosidase (*GBA*) cause Gaucher disease, an autosomal recessive lysosomal storage disorder. *GBA* mutation carriers have an elevated risk of Parkinson's disease (PD) compared with non-carriers and can experience an earlier age of onset and more aggressive disease. *GBA* analysis is technically challenging due to a related pseudogene and complex structural variations (SVs) that can occur at this locus. We have applied a recently developed nanopore sequencing method to analyze *GBA* variants in a New Zealand longitudinal cohort of PD, which includes extensive clinical assessment. We examined amplicons encompassing the coding region of *GBA* (8.9kb) from 229 PD cases using the GridION nanopore-sequencing platform. This revealed 23 mutations in 21 PD cases, which were confirmed via Sanger sequencing. We detected a strong PD risk allele p.N409S (rs76763715) in one case. Modest risk mutations included p.E365K (rs2230288) in 12 cases, and p.T408M (rs75548401) in seven cases, one of whom had p.E365K on the other allele. We also detected the possible risk alleles R78C (rs146774384) in one case, D179H (rs147138516) in one case, which occurred on the same haplotype as an E365K, and one novel pL335= (C>T) mutation that is predicted to impact splicing of *GBA* transcripts. This work confirmed the utility of nanopore sequencing as a high-throughput method to identify known and novel *GBA* mutations, and to assign precise haplotypes. Additionally, through clinical assessment, we provide support for the role of *GBA* mutations on disease aggression. This work may contribute our understanding of the effects of these mutations on disease presentation or progression, and to the development of more targeted treatments.

## Introduction

Several monogenic forms of Parkinson's disease (PD) and many genetic risk factors, which predispose to development of PD, are now known. One of the strongest genetic risk factors for PD is mutation of the gene known as *GBA*, which encodes the lysosomal enzyme glucocerebrosidase (GCase) [1]. GCase is an enzyme involved in glycolipid metabolism. Mutations in *GBA* lead to a marked decrease in GCase activity, which results in excessive accumulation of the GCase substrate, GlcCer in macrophages. If these mutations are homozygous, it causes the rare lysosomal storage disorder, Gaucher disease (GD) [2, 3]. Heterozygous *GBA* mutations were initially believed to be benign; however, parkinsonism symptoms were frequently reported in GD patients and their non-GD family members [4]. It is now clear that individuals with *GBA* mutations have a substantially elevated risk of developing PD compared with non-carriers [2, 5]. Furthermore, PD patients with *GBA* mutations have a noticeably earlier age of PD onset and a slightly increased risk of cognitive effects such as dementia [6]. A study by Odea et al. [7] also showed that PD patients carrying *GBA* mutations had a higher incidence of psychosis and delirium than PD patients who were non-carriers.

Over 300 different *GBA* mutations have been described, including point mutations (the majority of which are missense mutations), frameshift mutations, splice-site alterations, and a range of structural variants (SVs) [3]. However, the analysis of *GBA* is not straightforward. A nearby pseudogene called *GBAP1* shares 96% sequence homology with the coding region of the *GBA* gene, and care is required to prevent pseudogene contamination when analyzing *GBA* [8, 9]. The high homology and physical proximity between *GBA* and its pseudogene enables homologous recombination events, resulting in complex gene-pseudogene rearrangements [8], further compounding the difficulty of *GBA* analysis.

Although application of a long-PCR step targeted at *GBA* can effectively solve the pseudogene problem [9], challenges remain using short read sequencing with such amplicons, including the inability to precisely assign haplotypes. Long read sequencing approaches such as Oxford Nanopore Sequencing (Oxford Nanopore Technologies; ONT) and PacBio (Pacific Biosciences) offer new opportunities for analyzing complex genes such as *GBA*, and indeed a recent report described a procedure for carrying out genotyping and haplotyping on the MinION sequencer (ONT) [10]. We have further refined this method and described its application for the discovery of *GBA* variants in a New Zealand longitudinal cohort of PD patients.

## Materials and Methods

### Cohort description, recruitment and ethics

DNA for *GBA* genotyping was from a convenience sample (n = 229) of Parkinson's disease patients which were recruited through our movement disorders clinic for a range of PD-related research studies. Patients met the UK Brain Bank criteria for PD. Exclusion criteria included an indication of an atypical parkinsonian disorder, a history of moderate to severe head injury or stroke, a major psychiatric or medical illness in the previous six months, or poor English, which would prevent completion of neuropsychological testing. All participants had disease severity rated using the Movement Disorders Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS) and Hoehn and Yahr staging. Cognitive functioning was measured using a neuropsychological battery. Blood was collected from an arm vein into EDTA blood tubes. Whole blood was frozen until DNA extraction. All participants provided informed consent. The Southern Health and Disability Ethics committee (URA/11/08/042 and URB/09/08/037) approved the study protocols.

### DNA preparation and PCR

DNA from peripheral blood samples was extracted using a modified salting-out method followed by a phenol-chloroform purification step [11]. For each sample, a PCR that generated an 8.9 kilobase (kb) amplicon encompassing the entire coding region of *GBA* was performed according to Leija-Salazar et al. [10]. Agarose gel electrophoresis was used to verify the presence of the 8.9kb fragment for each sample.

Specific index sequences (ONT, Oxford, UK) were included in the 5′ end of both forward and reverse primers to enable sample barcoding and multiplexing in later steps. For each sample, 100ng of genomic DNA was amplified in the presence of 1 unit of Kapa Hifi Hotstart (Company, Location) ready mix and 0.3mM of each primer in a 50uL reaction. PCR consisted of 28 cycles of 95°C for 3 minutes, 98°C for 20 seconds, 60.9°C for 15 seconds, 72°C for 9 minutes, followed by a final extension at 72°C for 9 minutes. The primers including indexes are shown in Supplementary Table 1.

## Nanopore sequencing and validation of mutations

We established the nanopore method using several positive control samples carrying known *GBA* mutations (purchased from Coriell Institute, USA). Coriell IDs for these samples were NA01607, NA00372, NA10873, NA20270, NA02627 and NA20273.

*GBA* amplicons were purified using High-prep magnetic beads, modified using a protocol to remove fragments < 3-4kb [12]. To attach barcodes for sample multiplexing, we adapted a method described in the LSK-109 protocol (ONT, Oxford, UK). Reactions consisted of 140ng of the first round PCR product, 1uL PCR barcode (PCR Barcoding Expansion 1-96, EXP-PBC096), and 1 unit of Kapa Hifi Hotstart Ready Mix in a 50uL reaction. Thermocycler conditions consisted of 28 cycles of 95°C for 3 minutes, 98°C for 20 seconds, 62°C for 15 seconds, 72°C for 9 minutes, and one final extension of 72°C for 9 minutes. Following the barcoding reaction, another magnetic bead purification step was performed as described [12]. Ligation of the Nanopore adapters, library preparation, and flow cell priming was performed according to the LSK-109 protocol. Prepared libraries were run on R9.4 flowcells using the GridION platform (ONT) for between 24 and 48 hours with high-accuracy live base calling (Guppy v2.2.2 flip-flop).

Nested PCRs covering different regions on the 1:1000 diluted 8.9kb *GBA* PCR product were performed. Nested PCRs were carried out in 10µL reactions containing 1µL of a 1:10,000 dilution of the long PCR product in the presence of 1x Fisher TaqTi Buffer, 0.4mM dNTPs, 1.5mM $MgCl_2$, 0.05µL Fisher TAQ TI, 0.5mM of each primer and 1M betaine. Primers were designed using Geneious Prime® 2019.0.4 (https://www.geneious.com) to enrich for exons containing the mutation of interest (Supplementary Table 1). DNA was amplified in a touchdown PCR consisting of 15 cycles of 94°C for 15 seconds, 70°C for 15 seconds, and 72°C for 1 minute, with annealing temperature decrease of 1°C per cycle. This was followed by 20 cycles of 94°C for 15 seconds, 55°C for 15 seconds, and 72°C for 1 minute and a final extension at 72°C for 5 minutes. Sanger sequencing was subsequently performed on the nested PCR products (Supplementary Methods).

## Nanopore sequence data analysis

The software pipeline for processing nanopore FASTQ read files, and generating variant calls as VCF files was as follows. For reads with a quality-score (qscore) ≥ 7, we used Porechop

v0.2.4 (https://github.com/rrwick/Porechop) to remove sequencing adapters and subsequently demultiplex the reads. We did not use reads below this quality threshold for further analyses. Two different alignment tools, minimap2 v2.16 (r922) [13] and NGMLR v0.2.7 [14], were used to map reads to the *GBA* reference sequence (derived from Hg38) and generate the BAM alignment files. We simultaneously mapped the reads to the complete human reference genome (Hg38) to ensure that off target amplification had not occurred to a significant degree. We employed Qualimap v2.2 [15] and MinIONQC v1.4.1 [16] to obtain quality control metrics for the alignments and the nanopore reads, respectively. Read alignments to the human genome (Hg38) were visualized using Genome Ribbon [17].

SNPS and small indels were called using Nanopolish v0.11.1 [18] with ploidy2. Calling thresholds were set to 0.2 as recommended by the developer. The parameter --fix-homopolymers was included to prevent indel calls in homo-polymeric regions, which are generally false positives. To call larger SV's (>30bp changes), we used the program Sniffles v1.0.11 [14] with the default parameters. Alleles were phased for haplotyping using WhatsHap v0.18 [19] with the default parameters. File manipulation was performed using SAMtools v1.3.1 (https://github.com/samtools) and VCFtools v0.1.16 (https://vcftools.github.io).

To exclude false positives, we divided the Nanopolish quality score for the variant call by the total number of reads associated with the call. Adjusted quality scores >1.8 were considered true variants and <1.2 were considered false positives, the same threshold used by Leija-Salazar [10]. We further assessed the quality of the variant calls by manually checking the BAM files using IGV to ensure that the mutant reads were not a result of systematic error.

All variants were viewed on www.varsome.com, to assess their significance. Varsome provides a compilation of mutation data from gnomAD, dbSNP among others [20].

## Results

### Nanopore sequencing of GBA in NZPD

The nanopore sequencing method was applied to examine *GBA* amplicons derived from 229 DNA samples from the NZPD cohort. This was carried out using four libraries containing barcoded, multiplexed samples, on separate flow cells. All of the nanopore runs resulted in similar yields (~10-15 Gb with QC ≥ 7). The majority of these were the 8.9kb fragment of

interest (Supplementary Figure 1). For the first run ~60% of reads mapped to the *GBA* reference genome, and only a small fraction of the reads mapped to the pseudogene as indicated by Qualimap and visualization using Genome Ribbon [17] (Supplementary Figure 2). Coverage depth over *GBA* averaged >2000x across all samples as indicated by Qualimap, far in excess of the 100x coverage required for accurate *GBA* genotyping [10].

We validated the efficacy of the protocol for *GBA* genotyping by successfully identifying and calling eight different mutations, including six SNPs, one indel and one recombination mutant in positive control samples (Coriell Institute, USA). The alignment program minimap2 was preferred for variant calling due to its higher speed, although we simultaneously used NGMLR for alignment prior to SV detection, due to its compatibility with the SV caller Sniffles.

By only considering mutations with adjusted quality scores >1.8, all of the false positives were successfully filtered out, as verified by manually checking the BAM alignment files in Integrated Genomics Viewer (IGV) [21], and by Sanger sequencing validation of a subset of the predicted false positives.

## Potentially Pathogenic GBA mutations discovered

Using this approach, out of the 229 NZPD samples, we identified six different mutations of potential pathogenic significance over 21 different patients (Table 1); including two patients who each had two mutations

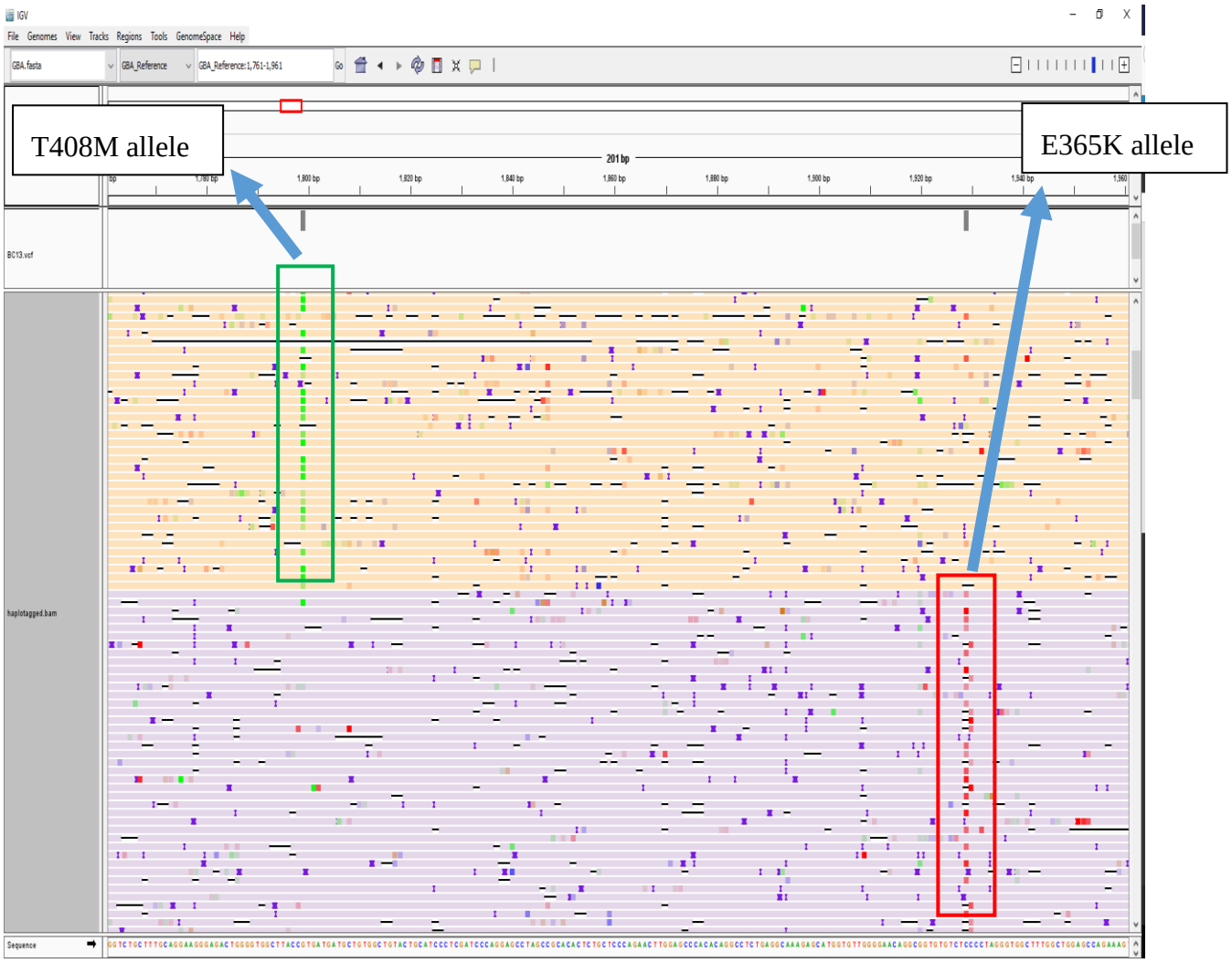
Table 1. Exon mutations detected in this cohort

| **Amino acid change** | **Old mutation nomenclature** | **Rs ID** | **Number of Samples** | **Number of Alleles** | **Cohort Allele Frequency** | **European non-Finnish Allele Frequency (GnoMAD)** |
|---|---|---|---|---|---|---|
| E365K | E326K | rs2230288 | 12 | 12 | 0.02620087 | 0.01234 |
| T408M | T369M | rs75548401 | 7 | 8 | 0.01746725 | 0.009459 |
| D179H | D140H | rs147138516 | 1 | 1 | 0.00218341 | 0.000272 |
| N409S | N370S | rs76763715 | 1 | 1 | 0.00218341 | 0.002045 |
| L335= | | Novel variant | 1 | 1 | 0.00218341 | N.A |
| R78C | R39C | rs146774384 | 1 | 1 | 0.00218341 | 0.000139 |
| **All Mutations** | | | **21** | 24 | 0.05240174 | |

The allele frequency of these mutations in the cohort is significantly elevated compared with the general population frequencies, as obtained from gnomAD. The mutation at the highest frequency in the cohort is E365K, with 12 patients detected. One patient in the cohort is heterozygous for both the E365K and the T408M mutation as indicated by allele phasing (Figure 1). We identified one patient with the D179H mutation who also carried E365K. In addition, one patient was found to have a novel variant (L335=) at a splice selection site of exon 8 (Supplementary Figure 4) [22]. We were unable to detect any SVs in the cohort using either Nanopolish or Sniffles.

A variety of other 'possibly benign' or 'uncertain significance' variants were detected in the cohort, including several novel intron SNPs and small deletions (Supplementary Table 2). All of the variants we detected were phased and visualized using inPHAP (Supplementary Figure 5) [23]. In total, we detected 74 different haplotypes within the cohort. 67% of the alleles were within four of the haplotypes and 47 of the haplotypes were represented by only a single allele (Supplementary Figure 6).

Figure 1. Example of *GBA* haplotypes generated by WhatsHap.




## Clinical phenotypes of mutation carriers

We noted a slightly decreased symptom (58.3) and diagnostic age (60.1) in the *GBA* mutation carrier group, compared with the non-mutation carriers (60.7 & 62.6). Furthermore, there was a notably higher prevalence of dementia (42.90%) in the *GBA* carrier group compared with the non-carrier group (20.20%) (Table 2).

Table 2. Demographic data for the PD patients in the New Zealand Parkinson's Disease cohort split by *GBA* mutation status.

| | n | Sex (% Male) | Education (yrs) | Follow-up duration (yrs) | Symptom Onset age (yrs) | Diagnosis age (yrs) | Disease duration (yrs) | Age at death (yrs) | Dementia (%) |
|---|---|---|---|---|---|---|---|---|---|
| *GBA* Mutation | 21 | 71 | $\bar{x}$ = 13.3 s.d =3.1 | $\bar{x}$ = 5.8 s.d =3.6 | $\bar{x}$ = 58.3 s.d =9.2 | $\bar{x}$ = 60.1 s.d =9.1 | $\bar{x}$ =15.2 s.d =6.9 | $\bar{x}$ = 76.30 s.d =8.4 | 42.90 |
| No *GBA* mutation | 208 | 67.30 | $\bar{x}$ = 12.8 s.d =2.6 | $\bar{x}$ = 5.1 s.d =3.6 | $\bar{x}$ = 60.7 s.d =8.5 | $\bar{x}$ = 62.6 s.d =8.5 | $\bar{x}$ = 13.7 s.d =5.8 | $\bar{x}$ =78.3 s.d =6.1 | 20.20 |

## Discussion

### Nanopore sequencing as a tool for GBA mutation discovery

We successfully validated the protocol of Leija-Salazar [10] as an effective method for genotyping of *GBA* across the 8.9kb amplicon. Furthermore, we highlight the capability of this protocol to phase variants, due to the single molecule reads generated by nanopore sequencing, thereby enabling haplotype establishment. Understanding haplotypes may reveal useful information about the origins and inheritance of *GBA* mutations.

We were unable to detect any SVs, despite previous studies reporting such variants in PD patients [3, 24]. The latter study reported four individuals presenting CNVs along the *GBA-GBAP1* region in a sporadic PD-cohort. While we could not detect any SVs, as with Leija-Salazar [10] we regularly noted that alignments by NGMLR were split between the gene and pseudogene. While these split reads could depict novel SVs, they tended to appear in a small number of reads meaning that chimeric molecules formed during the long PCR is a more likely explanation. This makes it difficult to resolve true complex SVs, a limitation of this protocol. Application of a novel PCR-free library generation method using CRISPR/ Cas9 capture of native DNA may be one possible solution to this limitation [25].

Another commonly noted limitation of nanopore sequencing is its inability to accurately resolve homopolymers and detect small insertions and deletions in these regions. Given this limitation, we simply filtered out these variants, meaning we may have overlooked some true variants. Continual hardware and software improvements in this system are likely to overcome this problem in due course.

## Profile of GBA mutations discovered in NZPD

In total, we detected mutations potentially relevant to the context of PD in 10.04% of patients in the cohort. This frequency is similar to what was anticipated based on previous studies [3].

### *Strong Risk allele: N409S*

Homologous N409S mutations are the most common variants associated with GD. There is good evidence that N409S increases the risk for PD significantly (OR = 7.27; 95% CI) [3, 4, 26]. However, the study by Ran et al. [26] showed that heterozygous N409S patients did not have an earlier age of onset, and did not differ significantly from other non-carrier PD patients.

### *Modest risk alleles: E365K and T408M*

The frequency of the E365K allele in our cohort is 0.0262, compared with the gnomAD frequency in the European (non-Finnish) population of 0.011. All 11 patients with E365K were heterozygotes.

E365K is not itself associated with GD, and only results in a mild loss-of-function enzyme. A recent meta-analysis indicates that there is a correlation between this mutation and PD in total populations, Asians, and Caucasians, respectively (OR = 1.99; 95% CI) [27]. That study also found that E365K predicted a more rapid progression of cognitive decline and motor symptom dysfunction.

The mutant allele frequency of T408M in the cohort is 0.0175 compared with the gnomAD European (non-Finnish) frequency of 0.009459. One of our six T408M positive patients is homozygous for this mutation, the rest are heterozygotes. T408M, like E365K, is not associated with GD, though it does result in decreased GCase enzyme activity [28]. A recent

meta-analysis reported that T408M is associated with an increased risk for PD (OR = 1.78; 95% CI) [29].
Additionally, both E365K and T408M mutations are frequently overrepresented in patients with REM-sleep behavior disorders (RBDs). Patients with RBDs have a significantly elevated risk of developing PD with dementia [2].

While these mutations are only modest risk factors for PD, due to the reduced enzyme function associated with both E365K and T408M [28, 29], patients carrying these mutations are good potential candidates for GCase-associated therapy.

### *Other Possible Risk alleles detected*

R78C has an allele frequency of 0.000139 in the gnomAD European (non-Finnish) population. While the clinical significance of this mutation remains unresolved, it has a "pathogenic" computational verdict based on seven predictions from DANN, GERP, dbNSFP, FATHMM, MetaLR, MetaSVM, MutationAssessor and MutationTaster (compared with three "benign" predictions from LRT, PROVEAN and SIFT). However, no papers have been published discussing this mutation in the context of GD or PD.

The gnomAD frequency for D179H is 0.00027 in European (non-Finnish) populations. The variant has been identified in GD patients [30], and UniProt classifies this variant as 'disease', but there are no studies regarding this mutation in a PD context. The patient carrying D179H also carries E365K, and haplotyping revealed that the two mutations exist on different alleles.

We also detected C>T; L335=, a novel synonymous mutation that it is predicted by Human Splice Finder to affect splice enhancer function (Supplementary Figure 4) [22]. The rarity of this mutation, combined with the splice predictions, means this mutation is a candidate for involvement in PD disease pathogenesis. However, further studies such as those involving analysis of mRNA splicing patterns in this patient would be necessary to confirm whether this variant may affect *GBA* splicing or function.

Ultimately, the lack of literature relating to these three possible risk alleles combined with only having a single data point for each, means that we cannot draw clear inferences regarding the effects of these mutations on clinical or developmental features of PD.

### Clinical implications

By comparing the clinical and demographic features of the patients between the *GBA* carrier and non-carrier group, we provide support to previous studies indicating that PD patients with *GBA* mutations have a noticeably earlier age of PD onset and a slightly increased risk of cognitive effects such as dementia [6]. Given the potential clinical effects of *GBA* mutations, and the interest in trialing drugs that target relevant pathways in *GBA* mutation positive patients, application of the nanopore sequencing method to detect a wide range of mutations at high throughput will have considerable utility.

## Conclusions

We have successfully validated ONT sequencing of *GBA* as an efficient and accurate method for detecting a wide range of variants in multiple patient samples, as well as resolving the haplotype of these variants. Applying this method to a sizeable New Zealand cohort of PD revealed the presence of multiple known *GBA* mutations, and a range of additional variants of unclear clinical significance.

These findings may contribute to a better understanding of the effects of these mutations on disease presentation or progression, and to development of more targeted treatments. This may be of importance as novel pharmaceuticals are actively under development to treat PD patients carrying *GBA* mutations.

## Acknowledgements

Support for this work came from the Jim and Mary Carney Charitable Trust (Whangarei, New Zealand), The McGee Fellowship (University of Otago, Christchurch), and The Helen Poole and Ian McDonald Memorial Summer Studentship (Canterbury Medical Research Foundation, Christchurch, New Zealand).

The Health Research Council and Lotteries Health provided funding for sample collection and storage. We would also like to acknowledge New Zealand Brain Research Institute Assistant Research Fellows Leslie Livingston, Sophie Grenfell, and Bob Young for cognitive and clinical assessments.

# References


1. O'Regan, G., et al., *Glucocerebrosidase Mutations in Parkinson Disease.* J Parkinsons Dis, 2017. **7**(3): p. 411-422.
2. Riboldi, G.M. and A.B. Di Fonzo, *GBA, Gaucher Disease, and Parkinson's Disease: From Genetic to Clinic to New Therapeutic Approaches.* Cells, 2019. **8**(4).
3. Sidransky, E. and G. Lopez, *The link between the GBA gene and parkinsonism.* The Lancet Neurology, 2012. **11**(11): p. 986-998.
4. Sidransky, E., et al., *Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease.* The New England journal of medicine, 2009. **361**(17): p. 1651-1661.
5. Halperin, A., D. Elstein, and A. Zimran, *Increased incidence of Parkinson disease among relatives of patients with Gaucher disease.* Blood Cells, Molecules, and Diseases, 2006.
6. Brockmann, K., et al., *GBA-associated Parkinson's disease: Reduced survival and more rapid progression in a prospective longitudinal study.* Movement Disorders, 2015. **30**(3): p. 407-411.
7. Oeda, T., et al., *Impact of glucocerebrosidase mutations on motor and nonmotor complications in Parkinson's disease.* Neurobiology of Aging, 2015.
8. Tayebi, N., et al., *Reciprocal and Nonreciprocal Recombination at the Glucocerebrosidase Gene Region: Implications for Complexity in Gaucher Disease*, in *Am J Hum Genet*. 2003. p. 519-34.
9. Jeong, S.Y., et al., *Identification of a novel recombinant mutation in Korean patients with Gaucher disease using a long-range PCR approach.* J Hum Genet, 2011. **56**(6): p. 469-71.
10. Leija-Salazar, 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): p. e564.
11. Miller, S.A., D.D. Dykes, and H.F. Polesky, *A simple salting out procedure for extracting DNA from human nucleated cells.* Nucleic acids research, 1988. **16**(3): p. 1215-1215.
12. Nagar, R. and B. Schewessinger, *DNA size selection (>3-4kb) and purification of DNA using an improved homemade SPRI beads solution.* 2018.
13. Li, H., *Minimap2: pairwise alignment for nucleotide sequences.* Bioinformatics, 2018. **34**(18): p. 3094-3100.
14. Sedlazeck, F.J., et al., *Accurate detection of complex structural variations using single-molecule sequencing.* Nature Methods, 2018. **15**(6): p. 461-468.
15. Okonechnikov, K., A. Conesa, and F. Garcia-Alcalde, *Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data.* Bioinformatics, 2016. **32**(2): p. 292-4.
16. Lanfear, R., et al., *MinIONQC: fast and simple quality control for MinION sequencing data.* Bioinformatics, 2018. **35**(3): p. 523-525.
17. Nattestad, M., C.-S. Chin, and M.C. Schatz, *Ribbon: Visualizing complex genome alignments and structural variation.* bioRxiv, 2016: p. 082123.
18. Loman, N.J., J. Quick, and J.T. Simpson, *A complete bacterial genome assembled de novo using only nanopore sequencing data.* Nature Methods, 2015. **12**: p. 733.
19. Martin, M., et al., *WhatsHap: fast and accurate read-based phasing.* bioRxiv, 2016: p. 085050.

20. Kopanos, C., et al., *VarSome: the human genomic variant search engine.* Bioinformatics, 2018. **35**(11): p. 1978-1980.

21. Robinson, J.T., et al., *Integrative genomics viewer.* Nature biotechnology, 2011. **29**(1): p. 24-26.

22. Desmet, F.-O., et al., *Human Splicing Finder: an online bioinformatics tool to predict splicing signals.* Nucleic acids research, 2009. **37**(9): p. e67-e67.

23. Jäger, G., A. Peltzer, and K. Nieselt, *inPHAP: interactive visualization of genotype and phased haplotype data.* BMC bioinformatics, 2014. **15**: p. 200-200.

24. Spataro, N., et al., *Detection of genomic rearrangements from targeted resequencing data in Parkinson's disease patients.* Movement Disorders, 2017. **32**(1): p. 165-169.

25. Gilpatrick, T., et al., *Targeted Nanopore Sequencing with Cas9 for studies of methylation, structural variants and mutations.* bioRxiv, 2019: p. 604173.

26. Ran, C., et al., *Strong association between glucocerebrosidase mutations and Parkinson's disease in Sweden.* Neurobiology of aging, 2016. **45**: p. 212.e5-212.e11.

27. Huang, Y., et al., *The Association between E326K of GBA and the Risk of Parkinson's Disease.* Parkinson's Disease, 2018. **2018**: p. 6.

28. Alcalay, R.N., et al., *Glucocerebrosidase activity in Parkinson's disease with and without GBA mutations.* Brain : a journal of neurology, 2015. **138**(Pt 9): p. 2648-2658.

29. Mallett, V., et al., *GBA p.T369M substitution in Parkinson disease: Polymorphism or association? A meta-analysis.* Neurol Genet, 2016. **2**(5): p. e104.

30. Montfort, M., et al., *Functional analysis of 13 GBA mutant alleles identified in Gaucher disease patients: Pathogenic changes and "modifier" polymorphisms.* Hum Mutat, 2004. **23**(6): p. 567-75.

## Figure Legend

Figure 1. Example of *GBA* haplotypes generated by WhatsHap. Phased variants for the E365K/T408M double mutant as visualized in IGV. The different haplotypes are displayed in yellow (top) and red (bottom). The top haplotype has the mutant allele at the T408M position (green box) and the 'normal' allele at the E365K position. The bottom haplotype has the mutant allele at the E365K position (red box) and the 'normal' allele at the T408M position.