Parkinsonism and Related Disorders 134 (2025) 107780

Contents lists available at ScienceDirect

Parkinsonism and Related Disorders

journal homepage: www.elsevier.com/locate/parkreldis

Short communication

LONG-NEXT: A new accurate and efficient NGS-based method for *GBA1* analysis in Parkinson disease

Giada Cuconato [a], Ilaria Palmieri [b], Marco Percetti [c], Serena Pagliarani [d], Sara Tenace [a], Marco J. Morelli [e], Ettore Zapparoli [e], Micol Avenali [f,g], Valerio Carelli [h,i], Patrizia Dentelli [j], Alessia Fiorentino [i], Andrea Gaudio [k], Claudia Ledda [l,m], Paola Mandich [k,o], Silvia Marino [p], Tiziana Martone [m,n], Raffaella Minardi [i], Paola Origone [k,o], Danara Ormanbekova [i], Barbara Pasini [j], Anna Scarabotto [b], Chiara Sorbera [p], Lucia Trevisan [k], the PARKNET Study Group, Alessio Di Fonzo [d], Enza Maria Valente [a,b,*], Edoardo Monfrini [c,d]

[a] *Department of Molecular Medicine, University of Pavia, Pavia, Italy*
[b] *Neurogenetics Research Center, IRCCS Mondino Foundation, Pavia, Italy*
[c] *Dino Ferrari Center, Neuroscience Section, Department of Pathophysiology and Transplantation, University of Milan, Milan, Italy*
[d] *IRCCS Foundation Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy*
[e] *Center for Omics Sciences, IRCCS San Raffaele Scientific Institute, Milan, Italy*
[f] *Department of Brain and Behavioural Sciences, University of Pavia, Pavia, Italy*
[g] *Parkinson Disease and Movement Disorders Unit, IRCCS Mondino Foundation, Pavia, Italy*
[h] *Department of Biomedical and Neuromotor Sciences, University of Bologna, Bologna, Italy*
[i] *IRCCS Istituto delle Scienze Neurologiche di Bologna, Programma di Neurogenetica, Bologna, Italy*
[j] *Department of Medical Science, University of Torino, Italy*
[k] *IRCCS Ospedale Policlinico San Martino, Genoa, Italy*
[l] *Department of Neurosciences Rita Levi Montalcini, University of Turin, Turin, Italy*
[m] *SC Neurologia 2U, AOU Città della Salute e della Scienza, Torino, Italy*
[n] *Medical Genetics Unit, Città della Salute e della Scienza University Hospital, 10126, Turin, Italy*
[o] *Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health (DINOGMI) University of Genoa, Genoa, Italy*
[p] *IRCCS Centro Neurolesi Bonino Pulejo, Messina, Italy*

ARTICLE INFO

Keywords:
Parkinson disease
GBA1 gene
Sanger sequencing
Short read NGS
LONG-NEXT

ABSTRACT

Introduction: GBA1 variants are the most common genetic risk factor for Parkinson disease (PD). Sequencing of *GBA1* on a large scale represents a burdensome task with currently adopted diagnostic techniques, namely Sanger sequencing and conventional short read next generation sequencing (sr-NGS). The high degree of sequence homology between *GBA1* and its pseudogene *GBA1LP* is the major driver behind this complexity, leading to false positive and false negative results.

We designed, optimized and validated LONG-NEXT, a new NGS-based strategy to streamline large scale *GBA1* sequencing.

Methods: LONG-NEXT relies on a specific long-range PCR, encompassing the whole *GBA1* gene, in one fragment (6.5 kb), followed by short-read NGS and a tailored bioinformatic pipeline masking the *GBA1LP* sequence on the reference genome.

Results: This protocol was optimized and tested on selected cases suspected of diagnostic mistakes by conventional testing (n = 13) and then validated on consecutively collected PD patients already screened either by Sanger sequencing (n = 101) or conventional sr-NGS (n = 294). LONG-NEXT reanalysis of 13 patients disclosed: 3 false positive cases due to mismapping of pseudogene reads on *GBA1*, 4 false homozygotes due to PCR-related allele dropout events, and 6 false negative cases, due to misalignment of *GBA1* reads against the pseudogene or PCR-related allele dropout events. The validation phase disclosed one additional false homozygote in the Sanger cohort, and one false negative result in the sr-NGS cohort.

* Corresponding author. Department of Molecular Medicine, University of Pavia, Via Mondino, 2, Pavia, 27100, Italy.
E-mail address: enzamaria.valente@unipv.it (E.M. Valente).

https://doi.org/10.1016/j.parkreldis.2025.107780
Received 10 December 2024; Received in revised form 18 February 2025; Accepted 9 March 2025
Available online 14 March 2025
1353-8020/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Conclusion: LONG-NEXT is a reliable, fast, cost-effective alternative for *GBA1* sequencing and may prove strategic in light of current genotype-based tailored therapies specifically targeting *GBA1*-PD patients.

1. Introduction

Biallelic pathogenic variants in the *GBA1* gene, encoding the lysosomal enzyme glucocerebrosidase (GCase), cause Gaucher disease (GD), the most common lysosomal storage disorder [1]. More recently, heterozygous variants in *GBA1* were found to represent the strongest genetic risk factor for Parkinson disease (PD), being detected in 5–15 % of all cases worldwide [2,3]. Of note, *GBA1* heterozygous genotype not only increases the risk of developing PD from 2 to >20 times compared to non-carriers, but also impacts on disease outcome, being associated with earlier age at onset, more rapid progression and higher occurrence of cognitive impairment and other non-motor symptoms [4,5]. Moreover, *GBA1* carrier status should be taken into account when leveraging therapeutic options in PD patients, and several clinical trials are ongoing, aimed at targeting GCase function in *GBA1*-PD patients [6]. For these reasons, molecular genetic testing of the *GBA1* gene is becoming increasingly relevant in clinical practice, and it is reasonable to hypothesize that such test may even become part of the diagnostic routine of PD at large.

The *GBA1* gene and its pseudogene (*GBA1LP*) are located in close proximity on chromosome 1q21.2, in a gene-rich genomic region which also includes the *MTX1* gene and the *MTX1P* pseudogene [7]. Since the difference in nucleotide sequence between *GBA1* and *GBA1LP* is very limited (96 % sequence homology), unequal homologous recombination can occur during meiosis, which can result in pseudogene-derived mutations, including gene conversions (nonreciprocal recombination) and gene-pseudogene fusions causing deletions or duplications (reciprocal recombination) [8].

Due to this reason, Sanger-based sequencing of the whole *GBA1* coding region (11 exons) is a burdensome and challenging task. The high degree of sequence homology between gene and pseudogene represents a major hurdle because of the high chance of amplifying and sequencing the pseudogene instead of the gene [8]. Moreover, PCR-related allele dropout (ADO) phenomena are not uncommon when amplifying *GBA1*, significantly reducing the sensitivity and specificity of this approach. The PCR specificity issue has been traditionally addressed by amplifying the *GBA1* region in three amplicons (encompassing exons 1–4, 5–7 and 8–11 respectively), using gene-specific primers specifically designed to anneal to those few nucleotides differentiating *GBA1* from its pseudogene [9]. However, this approach is costly and time-consuming, making it not amenable for screening large cohorts of PD patients.

The widespread advent of short-read NGS (sr-NGS) in diagnostic settings, allowing much faster and cheaper sequencing than Sanger, seemed to overcome this limit; yet NGS-based bioinformatic pipelines carry the risk of misaligning the sequencing reads of the gene to the pseudogene reference sequence and *vice versa*, especially when pseudogene-derived variants are involved [10]. Long-read NGS may address this issue representing an optimal solution for the problem of *GBA1* sequencing; however, its relatively high costs and unavailability in most diagnostic labs currently hinder its wide adoption in the clinical setting.

To address the need to provide fast and cost-effective analysis of *GBA1* at large scale, we designed and optimized a sr-NGS-based technology (LONG-NEXT) involving both "wet" and "*in silico*" analytical steps. We demonstrated the effectiveness of this approach in selected cases, and then validated it on consecutive samples previously tested for *GBA1* variants with a conventional technique (either Sanger sequencing or sr-NGS).

2. Materials and methods

2.1. Design of the study

The National Virtual Institute for PD, a consortium of Scientific Institutes for Medical Research and Care (IRCCS) focused on PD research funded by the Italian Ministry of Health, has undertaken PARKNET (see Supplementary Methods), an ambitious project to share knowledge and expertise, as well as clinical and genetic data of patients with PD among participating centres. Data-sharing agreements were signed by each institute involved. Written informed consent for research use of genetic data was obtained from each participant. The PARKNET genetic workgroup focused on the development of a new accurate and efficient approach to screen for *GBA1* mutations in large cohorts of PD patients. The sequencing protocol of LONG-NEXT (as described below) required one operator and allowed to genotype ~200 subjects per working week at an approximate reagent cost of 90€ per sample (~100$).

2.2. Subjects

LONG-NEXT was first tested on selected samples for which Sanger sequencing or sr-NGS performed over the past years had yielded doubtful results on *GBA1* genotype (optimization phase). For this phase, n = 13 cases were selected who fulfilled either of the following criteria:

1) genotype-phenotype discrepancy (e.g., homozygous *GBA1* pathogenic variants in PD patients without GD, or wild-type genotypes in patients with clinical features or family history suggestive of *GBA1*-PD);
2) ambiguous genetic result obtained with sr-NGS (e.g., *GBA1* variants with low variant allele frequency - VAF - or coverage depth anomalies).

For the subsequent validation phase, LONG-NEXT was assessed on two cohorts of PD patients already screened for *GBA1* mutations either by Sanger sequencing using a published protocol [9] (n = 101) or by conventional sr-NGS (n = 294).

PD diagnosis was made by specialists in movement disorders using the Movement Disorder Society clinical diagnostic criteria [11]. Written informed consent for genetic analyses for diagnostic and research purposes was obtained from all patients. All raw data are available from the corresponding author upon reasonable request.

2.3. LONG-NEXT protocol

The protocol consisted of the following steps briefly described below: i) long-range PCR; ii) library preparation and sequencing; iii) tailored bioinformatic analysis. Additional details on the methodology are provided in the Supplementary Methods.

2.3.1. Long-range PCR

The whole *GBA1* gene was amplified in a unique long-range PCR (6501bp) using GoTaq® Long PCR Master Mix (Promega, Madison, Wisconsin, US). Primers used were GBA1_6.5 kb_Forward: TCCTAAAGTTGTCACCCATACATG and GBA1_6.5 kb_Reverse: TAGTCACAGACAGCGTGTGAGC. The GBA1_6.5 kb_Forward was designed to anneal univocally to the 5′-UTR of *GBA1*, while the GBA_6.5 kb_Reverse primer was designed to anneal at both the 3′-UTR of *GBA1* and *GBA1LP*, to potentially detect also gene-pseudogene rearrangements. PCR amplification was performed with the following profile: 94 °C for 2 min;

30 cycles of 94 °C for 30 s, 62 °C for 30 s, and 65 °C for 7.30 min; 72 °C for 10 min. All PCR amplicons were purified before proceeding with library preparation.

2.3.2. Library preparation and sequencing

GBA1 long-range amplicons were used as input, and library preparation was performed using the Nextera XT DNA library preparation kit (Illumina, San Diego, California, US) following manufacturer's instructions. Normalized libraries were sequenced on an Illumina MiSeq instrument, using MiSeq Reagent Kits v2 (300 cycles), loading ~200 samples per run.

2.3.3. Tailored bioinformatic analysis

Bioinformatic analysis was conducted following GATK4 best practices (gatk4 4.6.1.0) [12], but employing the hg38 reference genome masked for the *GBA1LP* pseudogene sequence through "bedtools maskfasta'' command (bedtools v2.30.0) [13], allowing reads to align univocally against the *GBA1* reference sequence. This tailored pipeline displayed a significant increase in read depth and mapping quality compared to the standard approach (i.e. without pseudogene masking) and, more importantly, removed the risk of misaligning gene-specific sequences against the pseudogene, which is not an uncommon phenomenon when pseudogene-derived mutations are present (Supplementary Fig. 1). A minimal read depth of 500x was considered sufficient for analysis of variants.

2.4. Sanger sequencing and traditional sr-NGS

For Sanger sequencing, the *GBA1* gene was amplified through three gene-specific long-range PCRs as previously described [9] by using GoTaq® Long PCR (Promega), with some optimizations (see Supplementary Methods). Conventional sr-NGS was performed using distinct sr-NGS strategies (i.e. custom panel or whole-exome sequencing). Targeted enrichment approaches were either amplicon-based (119 samples) or hybrid capture (180 samples). Reads were aligned against the human reference genome (hg38) using BWA, variant calling was performed with GATK4 and variants were annotated using ANNOVAR.

3. Results

3.1. Optimization phase

LONG-NEXT was initially designed and optimized to verify selected cases of doubtful *GBA1* genotypes obtained with a conventional strategies (either Sanger sequencing or sr-NGS) (Supplementary Tables 1 and 2). For clarity purposes, we divided this paragraph in two subsections, each qualitatively comparing LONG-NEXT with either one of conventional approaches.

3.1.1. Sanger sequencing versus LONG-NEXT (n = 6)

LONG-NEXT disclosed heterozygous pathogenic *GBA1* variants in two PD patients presenting clinical features and family history suggestive of *GBA1*-PD but negative at Sanger sequencing (Supplementary Table 1). One (PD_12934) carried the c.721G>A (p.Gly241Arg) - G202R heterozygous variant (Supplementary Fig. 2), while the other (ME_PN13) harboured a heterozygous c.454+96_762-121del (Fig. 1A). Specularly, four PD patients without GD signs (20719-n.; 362VPF50; 3333/21; 147RPM70; Supplementary Table 1) who had previously been reported as homozygous carriers of *GBA1* variants [c.754T>A (p.Phe252Ile) - F213I; c.1223C>T (p.Thr408Met) - T369M; c.1093G>A (p.Glu365Lys) - E326K; c.1226A>G; (p.Asn409Ser) - N370S] were all correctly re-genotyped as heterozygous carriers by LONG-NEXT (Supplementary Fig. 3). The identified deletion could not be identified by Sanger sequencing due to removal of the annealing site of the forward primer used for amplification of the second *GBA1* fragment, and in fact it was already detectable by visual assessment of the long-range PCR

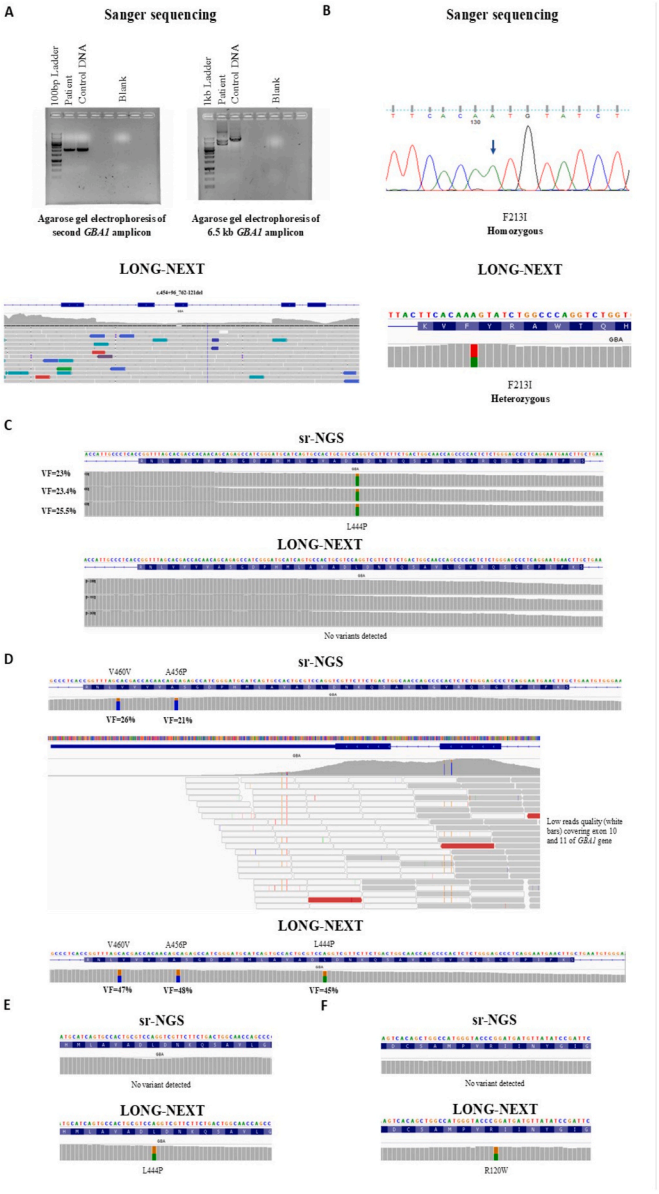

Fig. 1. Accuracy of LONG-NEXT versus conventional techniques in detecting *GBA1* pathogenic variants A-B: Sanger Sequencing versus LONG-NEXT. A) example of false negative case. Agarose gel electrophoresis comparison of *GBA1* second amplicon (CNV not detected) and 6.5 kb *GBA1* amplicon (CNV detected); IGV screenshot of the *.bam file showing the deletion (c.454+96_726-121del) encompassing exons 5 and 6 of *GBA1*, characterized by a decrease in coverage of the region of interest with well-defined deletion breakpoints; **B) example of false positive zygosity case**. Electropherogram of a patient carrier of F213I in homozygous state and IGV screenshot of the *.bam file reporting variant's correct zygosity. **C-F: sr-NGS versus LONG-NEXT. C) example of false positive cases**. IGV screenshot of three *.bam files from sr-NGS, "apparently" positive for L444P with unbalanced variant frequency, and from LONG-NEXT, showing absence of the variant; **D-F) example of false negative cases: D)** IGV screenshot of *.bam file from sr-NGS, showing V460V and A456P variants with unbalanced VF (26% and 21%) and poor coverage and quality reads mapping to *GBA1* exons 10 and 11 (white bars), and from LONG-NEXT correctly showing RecNcil variant; **E)** IGV screenshot of *.bam file from sr-NGS (*GBA1* exon 10) showing no variants, and from LONG-NEXT, showing the presence of L444P; **F)** IGV screenshot of *.bam file from sr-NGS (*GBA1* exon 5) showing no variants, and from LONG-NEXT, showing the presence of R120W.

product of the whole *GBA1* gene (6.5 kb) on agarose gel (Fig. 1A). All other false-negative and false-homozygous results were likely due to ADO events during the PCR amplification phase prior to Sanger sequencing.

3.1.2. Conventional sr-NGS versus LONG-NEXT (n = 7)

Three patients (190826; 221690; 223755; Supplementary Table 2), previously genotyped as heterozygous carriers of c.1448T>C (p.Leu483Pro) - L444P by sr-NGS, were found to be wild-type by LONG-NEXT (Fig. 1C). Interestingly, the VAF obtained by sr-NGS analysis was relatively low (patient 1 VAF = 23 %, patient 2 VAF = 23.4 %, patient 3 VAF = 25.5 %), but in all cases the variant had passed quality filters and was listed in the final annotated variant call format (*.vcf) files. Of note, c.1448T>C (p.Leu483Pro) - L444P is a constitutive variant of the *GBA1LP* pseudogene. Therefore, it is likely that the conventional NGS pipeline mapped the reads originating from the pseudogene on the genomic coordinates of the *GBA1* gene, erroneously calling the wild-type pseudogene sequence as a pathogenic *GBA1* genotype.

In another sample (3508-22; Supplementary Table 2), the sr-NGS pipeline revealed the synonymous c.1497G>C (p.Val499 =) - V460V variant with low VAF (26 %) in exon 10. From the direct visual inspection of the *.bam file, we observed an additional missense variant c.1483G>C (p.Ala495Pro) - A456P with very low VAF (21 %) in exon 10, which did not pass quality cutoff and was not listed in the final *.vcf file. Interestingly, the region encompassing *GBA1* exons 10 and 11 presented reduced coverage and low-quality mapping reads. This finding was suggestive of a likely misdetection of a recombinant allele. In particular, the RecNcil complex allele, originating from gene-pseudogene recombination from intron 10 to exon 11, results in the incorporation of the terminal segment of *GBA1LP* in the *GBA1* gene, including the three pseudogene variants c.1448T>C (p.Leu483Pro) - L444P, c.1483G>C (p.Ala495Pro) - A456P and c.1497G>C (p.Val499 =) - V460V. Re-sequencing of this case using LONG-NEXT displayed unequivocally that the patient was indeed a heterozygous carrier of the RecNcil mutation, with a balanced VAF for each variant and good quality sequencing reads (Fig. 1D). Another relevant example of misdetection of a *GBA1-GBA1LP* rearrangement was a sample (PoliMi_15231; Supplementary Table 2) whose sr-NGS analysis yielded a wild-type genotype. Interestingly, the *.bam file showed a decreased coverage across exon 9 of the gene, while increased coverage was observed in the corresponding region of the pseudogene, suggesting a misalignment of gene-derived sequencing reads to the pseudogene. In line with this hypothesis, LONG-NEXT clearly showed the presence of the c.1265_1319del [p.Leu422Profs*4 (Recdelta55)] mutation (Supplementary Fig. 4). Similarly, two samples (3434/22; 474VDM75; Supplementary Table 2) with wild-type genotypes at sr-NGS were subsequently found to carry heterozygous single nucleotide variants (SNV) with LONG-NEXT. Also in these cases, the detected variants [c.1448T>C (p.Leu483Pro) - L444P and c.475C>T (p.Arg159Trp) - R120W] belong to the pseudogene (Fig. 1E and F). Once more, the sr-NGS approach failed to call *GBA1* variants, and notably not only complex alleles, but also pseudogene-derived SNVs falling in regions of high homology between the gene and the pseudogene, as a result of an incorrect alignment process.

3.2. Validation phase

Prompted by these initial results, we validated LONG-NEXT comparing its diagnostic yield to either Sanger sequencing or sr-NGS in two consecutively collected cohorts of PD samples.

LONG-NEXT correctly detected all *GBA1* variants previously identified either by Sanger sequencing or sr-NGS (14/101 and 41/253, respectively). Notably, one sample in which Sanger detected the RecNcil complex variant (c.1448T>C (p.Leu483Pro) - L444P, c.1483G>C (p.Ala495Pro) - A456P and c.1497G>C (p.Val499 =) - V460V) in homozygosity was subsequently revealed to be heterozygous by LONG-NEXT (Supplementary Fig. 5, Supplementary Table 3). Heterozygosity was then confirmed by standard NGS approach. Moreover, one case testing negative at sr-NGS with amplicon enrichment technology was subsequently found with LONG-NEXT to carry the c.1448T>C (p.Leu483Pro) - L444P variant, which was then validated by Sanger sequencing (Supplementary Fig. 6, Supplementary Table 3). This provides additional evidence of sr-NGS-related false negative results caused by mis-mapping of reads in gene-pseudogene high homology regions.

4. Discussion

Sanger sequencing has been considered for many years the gold standard approach to search for *GBA1* variants, however this strategy is time-consuming and expensive, thus not suitable to face massive screenings. In the case of Sanger sequencing, the problem of sequence homology between gene and pseudogene was traditionally addressed by amplifying the *GBA1* region with three gene-specific amplicons [9]. However, this procedure is subjected to PCR-based biases, especially ADO events, possibly resulting in false negatives as well as false zygosity status. Furthermore, PCR-based bias could lead to the underdetection of structural variants such as large insertions and deletions/duplications, when they fall in-between the three fragments.

The advent of sr-NGS techniques has revolutionized genetic testing of PD patients, allowing the simultaneous analysis of many genes with reduced time and costs. Yet, the molecular analysis of the *GBA1* gene with such high-throughput technique remains challenging. The high degree of homology between *GBA1* and its pseudogene makes it difficult to properly map short reads to their correct reference sequence, especially in case of complex alleles derived by gene-pseudogene rearrangements. Sequence homology has an impact on both library preparation step and data analysis pipeline: commercial probes for library preparation are not specific for the *GBA1* gene, resulting in unspecific probe hybridization and consequent amplification of both gene and pseudogene, eventually affecting data analysis. Alignment tools are unable to determine if generated reads derive from the gene or the pseudogene, leading to decreased read depth, mapping quality and ultimately variant detection rate. The misalignment of pseudogene-derived reads against the gene causes false positive results, while the erroneous mapping of gene-derived reads against the pseudogene leads to false negative results, especially when pseudogene-derived variants are present. Interestingly, the accuracy of *GBA1* alignment seems to be affected by the choice of reference genome, with GRCh37/hg19 outperforming GRCh38/hg38 [14].

Several studies highlighted incongruencies in *GBA1* genotyping using sr-NGS, especially when referring to structural variants [14,15]. Recently, a specifically implemented algorithm was developed for resolving copy number variations (CNV) and complex alleles in *GBA1* locus from short-read WGS data (i.e., Gauchian) [15], but a subsequent study reported unresolved accuracy issues in this approach as well [16].

Long-read NGS sequencing may represent the optimal solution for the long-standing problem of *GBA1* sequencing; in particular, Long-Read Sequencing, using either Oxford Nanopore or Single Molecule Real-Time (SMRT) PacBio technologies, has been demonstrated to be a promising tool for sequencing complex genomic regions, although enrichment methods may be needed for the detection of *GBA1* recombinant class of variants [17–20]. Currently, this method is costly and limited in the diagnostic setting, narrowing its adoption for large-scale screening of PD patients [15].

For all these reasons, we propose an optimized sr-NGS-based approach which relies on a PCR-enriched step before library preparation, followed by a tailored data analysis pipeline, by improving a strategy already proposed in the context of GD [10]. First, we optimized a unique long-range PCR encompassing the whole *GBA1* gene (6501bp). Secondly, we used the entire human genome as a reference for mapping (except for masking *GBA1LP*), greatly enhancing the compatibility of the generated files with widely used data visualization and variant

interpretation tools. Unlike prior studies that focused on small sample sets, we validated LONG-NEXT in two independent cohorts, comparing the outcome versus both Sanger sequencing and conventional sr-NGS.

By presenting several paradigmatic cases, here we show that LONG-NEXT can achieve a better accuracy than conventional methods. Indeed, this approach successfully identified recombinant alleles, structural variants and pseudogene-derived variants, which were missed by standard approaches such as Sanger or conventional sr-NGS. In particular, sr-NGS is more prone to mistakes in detecting RecNcil and Rec(delta55) complex alleles, because reads are preferentially aligned to the pseudogene. As for recombinant alleles, pseudogene-derived variants localized in highly homologous regions may be missed. Conversely, several previously labelled *GBA1*-positive cases were reclassified as negative since reads belonging to the pseudogene had been erroneously aligned to *GBA1*, leading to false positive results. We also showed the intrinsic technical limitation of Sanger sequencing due to PCR biases (i.e., ADO) by successfully solving both false negative and false zygosity status with our approach.

Within the false negative group, we unravelled a gross multi-exonic deletion, completely missed by three-fragment Sanger sequencing analysis, demonstrating the capability of LONG-NEXT to detect also this type of rearrangements.

Remarkably, discordant cases emerged also from the validation cohorts. We disclosed one false zygosity status in a Sanger sequenced sample, due to an ADO event, as well as a false negative result obtained with sr-NGS, that missed the pseudogene-derived variant L444P. Notably, the same L444P variant has been correctly identified by sr-NGS in additional 4 cases, demonstrating variable accuracy in detecting pseudogene-derived variants across different sr-NGS methodologies and bioinformatic pipelines. In particular, hybrid capture technologies appear generally superior to amplicon-based approaches. Moreover, lower quality filter thresholds for variant calling may increase sensitivity in detecting such variants.

LONG-NEXT method exploits all the advantages of parallel massive sequencing, being time- and cost-effective in comparison to classical Sanger sequencing, but also overcoming the main limitations of standard sr-NGS. In light of current research evaluating the impact of *GBA1* variants on PD prognosis [5], clinical outcome and therapeutic choices [21], the availability of an optimized approach for large-scale *GBA1* sequencing in the diagnostics setting becomes critical.

Despite the significant advantages of LONG-NEXT in terms of reliability and speed of analysis, we acknowledge some limitations of this approach. The first limitation regards the possibility of ADO. While these events are less likely in comparison to Sanger sequencing approach (only one amplicon is generated instead of three), they cannot be completely excluded. In fact, rare variants may fall in the genomic region where the primers anneal resulting in ADO event. However, when designing primers for the long-range PCR, we carefully checked for SNPs using a tool (https://genetools.org/SNPCheck/snpcheck.htm). We eventually selected a pair of primers harbouring a single rare SNP only in the reverse sequence (0.000014 % gnomAD genomes), thus minimizing the probability of ADO. Furthermore, large complex recombinations may not be completely included in the region amplified by the long-range PCR [22]. To overcome this issue, we are improving the protocol by standardizing coverage across a given number of samples for each run, obtaining a medium read-depth score. Hence, a significant downward deviation from the standard value could suggest ADO and therefore a complex recombination event, which will need to be validated with a complementary method, such as multiplex ligation-dependent probe amplification (MLPA). A second limitation is that the 5′UTR region is not included in the long-range PCR. This also occurs with the conventional 3-fragment *GBA1* Sanger analysis, while sr-NGS usually targets this segment, which theoretically may contain rare pathogenic variants. We initially attempted a long-range PCR covering the 5′UTR (~12 kb, data not shown), however the success rate of this PCR was scarce. Since our approach aimed at large scale analysis and no known pathogenic variants have been identified in 5′UTR of *GBA1* gene (https://pdgenetics.shinyapps.io/gba1browser/), we opted to exclude the 12 kb long-range PCR in favour of the more efficient 6.5 kb amplicon. Finally, to be a time and cost-efficient approach, the minimum number of samples per run should be ~100. Those numbers may be difficult to reach for a single genetic centre and this might have implications on diagnostic turn-around time. In this view, the development of collaborative networks such as PARKNET, where few specialized hub centres can simultaneously process samples received by all spokes, could represent a solution to this challenge. It is also worth mentioning that libraries of different amplicons can be merged and simultaneously sequenced into the same flow cell, making LONG-NEXT also suitable for laboratories processing lower sample numbers.

In conclusion, LONG-NEXT is a reliable, fast, and cost-effective alternative for large-scale sequencing of the *GBA1* gene. It is broadly applicable in clinical diagnostic laboratories, showing high accuracy in detecting *GBA1* variants, including those that may be missed with alternative techniques. LONG-NEXT represents a strategic solution to the unmet need of *GBA1* genotyping in large cohorts of PD patients, to offer tailored treatments and experimental therapies specifically targeting this frequent genetic subgroup.

CRediT authorship contribution statement

Giada Cuconato: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation. **Ilaria Palmieri:** Writing – review & editing, Investigation. **Marco Percetti:** Writing – review & editing, Resources. **Serena Pagliarani:** Writing – review & editing, Resources. **Sara Tenace:** Writing – review & editing, Resources. **Marco J. Morelli:** Writing – review & editing, Investigation. **Ettore Zapparoli:** Writing – review & editing, Investigation. **Micol Avenali:** Writing – review & editing, Resources. **Valerio Carelli:** Writing – review & editing, Resources. **Patrizia Dentelli:** Writing – review & editing, Resources. **Alessia Fiorentino:** Writing – review & editing, Investigation. **Andrea Gaudio:** Writing – review & editing, Investigation. **Claudia Ledda:** Writing – review & editing, Investigation. **Paola Mandich:** Writing – review & editing, Resources. **Silvia Marino:** Writing – review & editing, Resources. **Tiziana Martone:** Writing – review & editing, Resources. **Raffaella Minardi:** Writing – review & editing, Investigation. **Paola Origone:** Writing – review & editing, Resources. **Danara Ormanbekova:** Writing – review & editing, Investigation. **Barbara Pasini:** Writing – review & editing, Resources. **Anna Scarabotto:** Writing – review & editing, Investigation. **Chiara Sorbera:** Writing – review & editing, Resources. **Lucia Trevisan:** Writing – review & editing, Resources. **Alessio Di Fonzo:** Writing – review & editing, Supervision, Resources. **Enza Maria Valente:** Writing – review & editing, Supervision, Funding acquisition. **Edoardo Monfrini:** Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Conceptualization.

Data availability statement

The datasets generated during the current study are available from the corresponding author upon reasonable request.

Funding

This work was supported with funds from the Italian Ministry of Health (RC2022-2024, RC Reti).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.parkreldis.2025.107780.

References

[1] J. Stirnemann, et al., A review of gaucher disease pathophysiology, clinical presentation and treatments, Int. J. Mol. Sci. 18 (2) (Feb. 2017), https://doi.org/10.3390/ijms18020441.

[2] E. Sidransky, et al., Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease, N. Engl. J. Med. 361 (17) (Oct. 2009) 1651–1661, https://doi.org/10.1056/NEJMoa0901281.

[3] G.M. Riboldi, A.B. Di Fonzo, 'GBA, gaucher disease, and Parkinson's disease: from genetic to clinic to new therapeutic approaches.', Cells 8 (4) (Apr. 2019) https://doi.org/10.3390/cells8040364.

[4] Z. Gan-Or, et al., Differential effects of severe vs mild GBA mutations on Parkinson disease, Neurology 84 (9) (Mar. 2015) 880–887, https://doi.org/10.1212/WNL.0000000000001315.

[5] S. Petrucci, et al., GBA-related Parkinson's disease: dissection of genotype-phenotype correlates in a large Italian cohort, Mov. Disord. 35 (11) (Nov. 2020) 2106–2111, https://doi.org/10.1002/mds.28195.

[6] F. Cavallieri, et al., Recent advances in the treatment of genetic forms of Parkinson's disease: hype or hope? Cells 12 (5) (Feb. 2023) https://doi.org/10.3390/cells12050764.

[7] A. Velayati, M.A. Knight, B.K. Stubblefield, E. Sidransky, N. Tayebi, Identification of recombinant alleles using quantitative real-time PCR implications for Gaucher disease, J. Mol. Diagn. 13 (4) (Jul. 2011) 401–405, https://doi.org/10.1016/j.jmoldx.2011.02.005.

[8] N. Tayebi, et al., Reciprocal and nonreciprocal recombination at the glucocerebrosidase gene region: implications for complexity in Gaucher disease, Am. J. Hum. Genet. 72 (3) (Mar. 2003) 519–534, https://doi.org/10.1086/367850.

[9] D.L. Stone, N. Tayebi, E. Orvisky, B. Stubblefield, V. Madike, E. Sidransky, Glucocerebrosidase gene mutations in patients with type 2 Gaucher disease, Hum. Mutat. 15 (2) (2000) 181–188, https://doi.org/10.1002/(SICI)1098-1004(200002)15:2<181::AID-HUMU7>3.0.CO;2-S.

[10] S. Zampieri, S. Cattarossi, B. Bembi, A. Dardis, GBA analysis in next-generation era: pitfalls, challenges, and possible solutions, J. Mol. Diagn. 19 (5) (Sep. 2017) 733–741, https://doi.org/10.1016/j.jmoldx.2017.05.005.

[11] R.B. Postuma, et al., MDS clinical diagnostic criteria for Parkinson's disease, Mov. Disord. 30 (12) (Oct. 2015) 1591–1601, https://doi.org/10.1002/mds.26424.

[12] G.A. Van der Auwera, et al., From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline, Curr Protoc Bioinformatics 43 (1110) (2013) 11.10.1–11.10.33, https://doi.org/10.1002/0471250953.bi1110s43.

[13] A.R. Quinlan, I.M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features, Bioinformatics 26 (6) (Mar. 2010) 841–842, https://doi.org/10.1093/bioinformatics/btq033.

[14] K. Orimo, et al., Association study of GBA1 variants with MSA based on comprehensive sequence analysis -Pitfalls in short-read sequence analysis depending on the human reference genome-, J. Hum. Genet. 69 (12) (Dec. 2024) 613–621, https://doi.org/10.1038/s10038-024-01266-1.

[15] M. Toffoli, et al., Comprehensive short and long read sequencing analysis for the Gaucher and Parkinson's disease-associated GBA gene, Commun. Biol. 5 (1) (Jul. 2022) 670, https://doi.org/10.1038/s42003-022-03610-7.

[16] N. Tayebi, J. Lichtenberg, E. Hertz, E. Sidransky, Is Gauchian genotyping of GBA1 variants reliable? medRxiv (Oct. 2023) https://doi.org/10.1101/2023.10.26.23297627.

[17] S. Pachchek, et al., Accurate long-read sequencing identified GBA1 as major risk factor in the Luxembourgish Parkinson's study, NPJ Parkinsons Dis 9 (1) (Nov. 2023) 156, https://doi.org/10.1038/s41531-023-00595-w.

[18] M. Leija-Salazar, et al., Evaluation of the detection of GBA missense mutations and other variants using the Oxford Nanopore MinION, Mol Genet Genomic Med 7 (3) (Mar. 2019) e564, https://doi.org/10.1002/mgg3.564.

[19] G.I. Drelichman, et al., Long-read single molecule real-time (SMRT) sequencing of GBA1 locus in Gaucher disease national cohort from Argentina reveals high frequency of complex allele underlying severe skeletal phenotypes: collaborative study from the Argentine Group for Diagnosis and Treatment of Gaucher Disease, Mol Genet Metab Rep 29 (Dec. 2021) 100820, https://doi.org/10.1016/j.ymgmr.2021.100820.

[20] C. Gabbert, et al., GBA1 in Parkinson's disease: variant detection and pathogenicity scoring matters, BMC Genom. 24 (1) (Jun. 2023) 322, https://doi.org/10.1186/s12864-023-09417-y.

[21] M. Avenali, et al., Are patients with GBA-Parkinson disease good candidates for deep brain stimulation? A longitudinal multicentric study on a large Italian cohort, J. Neurol. Neurosurg. Psychiatry (Oct. 2023), https://doi.org/10.1136/jnnp-2023-332387.

[22] A. Velayati, M.A. Knight, B.K. Stubblefield, E. Sidransky, N. Tayebi, Identification of recombinant alleles using quantitative real-time PCR implications for Gaucher disease, J. Mol. Diagn. 13 (4) (Jul. 2011) 401–405, https://doi.org/10.1016/j.jmoldx.2011.02.005.