Long-Read Sequencing Resolves a Complex Structural Variant in *PRKN* Parkinson's Disease

Kensuke Daida, MD, PhD,[1,2,3] Manabu Funayama, PhD,[3,4] Kimberley J. Billingsley, PhD,[5] Laksh Malik, MFS,[2] Abigail Miano-Burkhardt, BSc,[5] Hampton L. Leonard, MS,[2,5,6,7] Mary B. Makarious, BSc,[5,8,9] Hirotaka Iwaki, MD, PhD,[2,6] Jinhui Ding, PhD,[10] J. Raphael Gibbs, PhD,[10] Mayu Ishiguro, MD,[3] Hiroyo Yoshino, PhD,[4] Kotaro Ogaki, MD, PhD,[3] Genko Oyama, MD, PhD,[3] Kenya Nishioka, MD, PhD,[11] Risa Nonaka, PhD,[3,12,13] Wado Akamatsu, MD, PhD,[13] Cornelis Blauwendraat, PhD,[1,2*] and Nobutaka Hattori, MD, PhD[3,4,14*]

[1]Integrative Neurogenomics Unit, Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland, USA
[2]Center for Alzheimer's and Related Dementias (CARD), National Institute on Aging and National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland, USA
[3]Department of Neurology, Faculty of Medicine, Juntendo University, Tokyo, Japan
[4]Research Institute for Diseases of Old Age, Graduate School of Medicine, Juntendo University, Tokyo, Japan
[5]Molecular Genetics Section, Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland, USA
[6]Data Tecnica International LLC, Washington, DC, USA
[7]German Center for Neurodegenerative Diseases (DZNE), Tübingen, Germany
[8]Department of Clinical and Movement Neurosciences, UCL Queen Square Institute of Neurology, London, United Kingdom
[9]UCL Movement Disorders Centre, University College London, London, United Kingdom
[10]Computational Biology Group, Laboratory of Neurogenetics, National Institute on Aging, NIH, PorterNeuroscience Research Center, Bethesda, Maryland, USA
[11]Department of Neurology, Juntendo Tokyo Koto Geriatric Medical Center, Tokyo, Japan
[12]Department of Clinical Data of Parkinson's Disease, Graduate School of Medicine, Juntendo University, Tokyo, Japan
[13]Center for Genomic and Regenerative Medicine, Graduate School of Medicine, Juntendo University, Tokyo, Japan
[14]Neurodegenerative Disorders Collaborative Laboratory, RIKEN Center for Brain Science, Saitama, Japan

ABSTRACT: Background: *Parkin RBR E3 ubiquitin-protein ligase* (*PRKN*) mutations are the most common cause of young onset and autosomal recessive Parkinson's disease (PD). *PRKN* is located in FRA6E, which is one of the common fragile sites in the human genome, making this region prone to structural variants. However, complex structural variants such as inversions of *PRKN* are seldom reported, suggesting that there are potentially unrevealed complex pathogenic *PRKN* structural variants.

Objectives: To identify complex structural variants in *PRKN* using long-read sequencing.
Methods: We investigated the genetic cause of monozygotic twins presenting with a young onset dystonia-parkinsonism using targeted sequencing, whole exome sequencing, multiple ligation probe amplification, and long-read sequencing. We assessed the presence and frequency of complex inversions overlapping *PRKN* using whole-genome sequencing data of Accelerating Medicines

Correspondence to: Dr. Cornelis Blauwendraat, Integrative Genomics Unit, Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, 35 Convent Drive, Bethesda, MD 20892, USA; E-mail: cornelis.blauwendraat@nih.gov; Prof. Nobutaka Hattori, Department of Neurology, Faculty of Medicine, Juntendo University, 2-1-1 Hongo, Bunkyo-ku, Tokyo 113-8421, Japan; E-mail: nhattori@juntendo.ac.jp

Partnership Parkinson's disease (AMP-PD) and United Kingdom (UK)-Biobank datasets.

Results: Multiple ligation probe amplification identified a heterozygous exon three deletion in *PRKN* and long-read sequencing identified a large novel inversion spanning over 7 Mb, including a large part of the coding DNA sequence of *PRKN*. We could diagnose the affected subjects as compound heterozygous carriers of *PRKN*. We analyzed whole genome sequencing data of 43,538 participants of the UK-Biobank and 4941 participants of the AMP-PD datasets. Nine inversions in the UK-Biobank and two in AMP PD were identified and were considered potentially damaging and likely to affect *PRKN* expression.

Conclusions: This is the first report describing a large 7 Mb inversion involving breakpoints outside of *PRKN*. This study highlights the importance of using long-read sequencing for structural variant analysis in unresolved young-onset PD cases. © 2023 The Authors. *Movement Disorders* published by Wiley Periodicals LLC on behalf of International Parkinson and Movement Disorder Society. This article has been contributed to by U.S. Government employees and their work is in the public domain in the USA.

Key Words: PARK2; long-read sequencing; structural variant; Parkinson's disease; inversion

Parkinson's disease (PD) is the second most common neurodegenerative disease next to Alzheimer's disease, developing with symptoms such as bradykinesia, resting tremor, rigidity, and postural instability.[1] PD is typically divided into two categories: young-onset PD and late-onset PD. The cutoff age that defines young onset PD has been ambiguous, but it has recently been defined by the Movement Disorder Task Force as any age at onset before the age of 50.[2] Genetics plays a vital role in young onset PD and the *parkin RBR E3 ubiquitin-protein ligase* (*PRKN*) gene is the most frequent causative gene in young onset PD and autosomal recessive PD,[3] which were initially identified in a Japanese family.[4] *PRKN* encodes the parkin protein, which is an E3 ubiquitin-protein ligase that maintains mitochondrial function by removing dysfunctional mitochondria through autophagy. This process heavily involves *PINK1*, where *PINK1* mutations are the second most common cause of young-onset PD and autosomal recessive PD.[5]

PRKN is located on chromosome 6q25.2-27, containing 12 coding exons and spanning a total of 1.3 Mb. The region of *PRKN* is within FRA6E, which is the third most frequently observed common fragile site of the human genome.[6,7] Common fragile sites are specific loci that preferentially exhibit gaps and breaks on metaphase chromosomes following partial inhibition of DNA synthesis.[8] The central core of FRA6E is located in exon 3 to 8 of *PRKN*, which is the known mutation hot spot of *PRKN*.[7]

To date, over a 100 pathogenic *PRKN* mutations including point mutation and structural variants (SVs) are reported in the Movement Disorder Society Genetic mutation database (https://www.mdsgene.org/). Biallelic *PRKN* mutation accounts for an estimated 4.3% and 8% to 15% of sporadic and familial young onset PD.[9,10] The role of heterozygous *PRKN* variants in PD is controversial, where some reports show increased risk of carrying a single damaging variant and others report no effect.[11-16] It is worth noting that the reported associations between heterozygous *PRKN*

and PD in those studies may be influenced by a potential second unrevealed and complex to identify mutation that affects the function of *PRKN*.

Biallelic *PRKN* patients typically present young or juvenile-onset parkinsonism with levodopa responsiveness, dystonia, dyskinesia, and motor fluctuations, but without cognitive decline, autonomic symptoms, and psychotic symptoms.[17] The pathology of the brain with *PRKN* variants is characterized by neuronal loss in substantia nigra pars compacta often without Lewy body.[18]

Long-read sequencing has been an emerging sequencing technique where reads longer than 10,000 bp can reliably be sequenced, which is >30 times larger than conventional short-read sequencing.[19] Not only do longer reads (10 kb +) improve de novo assembly and mapping, but they are better able to detect SVs because they can span repetitive or other complex regions of the genome. For identifying SVs in PD specifically, a recent study that compared matched long and short sequencing read data from PD cases highlighted that most SVs in the human genome are likely undetectable with short-read data alone (~84%).[20] Long-read sequencing is a powerful tool for identifying potential causal variants that were previously undetectable using other sequencing technologies.

Here, we describe a *PRKN* family with two affected monozygotic twins presented clinically with dystonia parkinsonism. No other family members had been diagnosed with PD or any other neurological diseases. Initial genetic testing identified a single *PRKN* mutation despite showing a clear typical *PRKN* clinical presentation. After extensive follow-up, using long-read sequencing we identified an additional complex SV at the *PRKN* locus explaining the *PRKN*-PD phenotype.

Methods

Study Participants

The study was approved by the ethics committee of Juntendo University, Tokyo, Japan, and all participants provided written informed consent to participate in the

genetic research described in this study. PD was clinically diagnosed according to standard clinical criteria.[21,22] DNA was extracted from peripheral blood by the standard protocol using QIAamp DNA Blood Maxi Kit (QIAGEN, Venlo, Netherlands).

Targeted Panel Sequencing Using Short Read Whole Exome Sequencing

Targeted panel sequencing and whole exome sequencing (WES) were performed in II-3, III-1, and III-2 (Fig. 1A). The method of targeted panel sequencing for PD-related genes has been previously reported.[23] Libraries for WES were prepared with SureSelect Human All Exon V6 kit (Agilent Technologies Santa Clara, CA). Libraries were sequenced using the Illumina Hiseq1500 (Illumina, San Diego, CA). Picard was used to mark duplicates, and variants were called according to Genome Analysis Toolkit version 4.1.3.0 pipeline.[24] Annotation was conducted by Annovar.[25] Variants were filtered according to the following criteria: base quality score, location in exons or splice sites, and the allele frequency (gnomAD) smaller than 0.001.[26]

Multiplex Ligation-Dependent Probe Amplification

Copy number variants (CNVs) in *PRKN* and *SNCA* were analyzed using multiplex ligation-dependent probe amplification (MLPA) with selective adaptor ligation, selective amplification (SALSA) MLPA Probemix P051/P052 Parkinson probe mix (MRC-Holland, Amsterdam, the Netherlands). MLPA experiments were done according to the manufacturer's instructions.

Oxford Nanopore Technologies Long-Read Sequencing

We used the DNA prepared for the short-read sequencing as starting material for the long-read sequencing. Sequencing was prepared according to our protocol reported previously.[27] In brief, DNA samples were sized using the Femto Pulse (Agilent Technologies Santa Clara, CA). DNA underwent through a size selection step using the Circulomics Short Read Eliminator Kit (SS-100-101-01) to remove fragments up to 25 kb. Libraries were prepared using the Kit V14 Ligation sequencing kit from ONT and sequenced using PromethION for 72 hours on an R10.4.1 flow cell (Oxford Nanopore Technologies, Oxford, United Kingdom [UK]). Base calling was performed by Guppy version 6.3.8,[28] and Winnowmap version 2.0.3[29] was used to map the reads to the GRCh38 reference genome. Sniffles version 2.0.7[30] was used for calling SVs (using tandem-repeats option). SVs were annotated by AnnotSV version 3.1.1.[31]

Inversion Breakpoint Region Sequence Analysis

To confirm the breakpoints of the inversion that were identified by long-read sequencing, we amplified the breakpoint regions using polymerase chain reaction (PCR) by primers specifically designed by Primer 3 to detect the inversion (Supplementary Table S1). To consider how the breakpoints occur, we performed breakpoint region sequence analysis as described previously.[32,33] We obtained the sequence of 100 bp upstream and downstream of the breakpoints from the UCSC genome browser. (https://genome.ucsc.edu/index.html). Repeat information was obtained by RepeatMasker (Smit, A.F.A., Hubley, R., and Green, P. RepeatMasker Open-4.0. 2013–2015, http://www.repeatmasker.org). Sequence similarity around the breakpoints was checked using EMBOSS Needle.[34] We also used palindrome software to identify sequences with potential to cause stem loops (https://emboss.bioinformatics.nl/cgi-bin/emboss/palindrome).

Illumina Short Read Whole Genome Sequencing

To replicate the findings in long-read sequencing, we conducted short-read whole genome sequencing (WGS) to III-2. WGS was performed by library preparation using the VAHTS Universal Pro DNA Library Prep Kit from Illumina and sequencing was performed using an Illumina Novaseq 6000 sequencer at 150 bp paired-end sequencing. The estimated data generated per sample was >90 Gb resulting in >30 coverage. Reads were aligned by Burrows-Wheeler Alignment tool version 0.7.17.[35] Short-read WGS was analyzed in the same framework as the short-read WES, which is written above. SVs were called using Manta version 1.6.0.[36]

Replication Using Short Read WGS Datasets

As the inversion was also identified after targeted re-analysis of the short-read sequencing, we used the UK Biobank short-read WGS data, Accelerating Medicines Partnership PD (AMP-PD) cohort to examine the frequency of the inversions of *PRKN* in controls and cases. AMP-PD cohort is a dataset of short-read WGS, which are described previously.[16] We called the SVs using Manta[36] and used the output of diploidSV.vcf.gz. We used the Manta-called scored SV and indel candidates in UK Biobank (Field ID 23350). Using both datasets, the variants that matched the following criteria remained for further analysis: (1) variant type is inversion; (2) the inversion affects the transcript of *PRKN*; and (3) the size of the SV is within 50 bp to 10 Mb. The filtered variants were confirmed visually using integrative genomics viewer (IGV) to filter false-positive inversions (called as inversions for several chromosomes).[37]
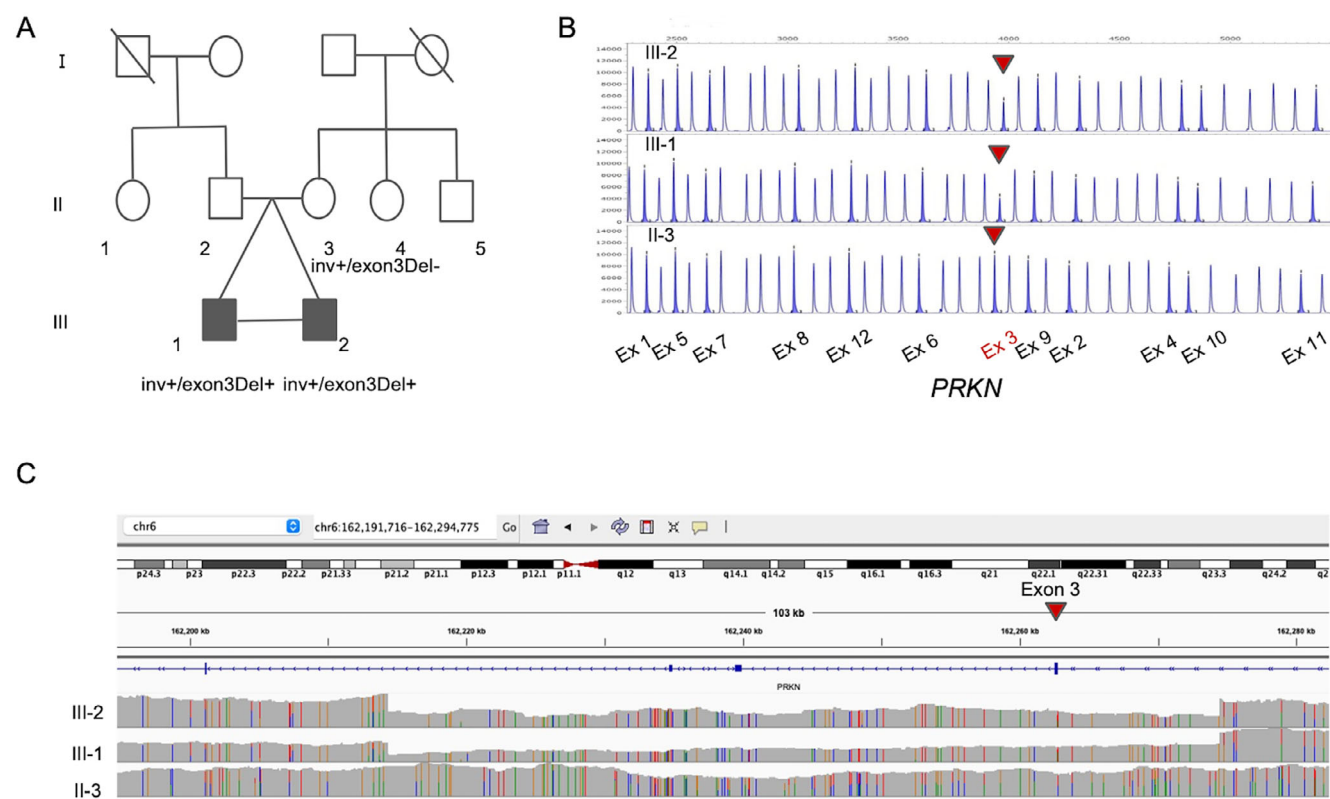
FIG. 1. Family pedigree and genetic test results of the twin's family. (**A**) Family pedigree with *Parkin RBR E3 ubiquitin-protein ligase* (*PRKN*) inv/exon 3 deletion. Squares, men; circles, women; oblique lines, deceased; black squares, clinically diagnosed with Parkinson's disease (PD). (**B**) Multiplex ligation-dependent probe amplification (MLPA) result for *PRKN* exons. (**C**) Screenshot from integrative genomics viewer (IGV) presenting with exon 3 deletion in *PRKN*. inv; inversion.

Data Sharing

UK Biobank data is available on application at the UK Biobank website (https://www.ukbiobank.ac.uk/). AMP-PD data is available on application at the AMP-PD website (https://amp-pd.org/).

Results

Here, we report monozygotic twins presented with a young onset dystonia parkinsonism phenotype (Fig. 1A). They were born without any problems during the pregnancy. Their family is of Japanese ancestry and has no reported consanguinity. None of their family members including their parents had been diagnosed with PD or any neurological diseases (Fig. 1A). The clinical symptoms of the affected twins at the age of 31 are summarized in Supplementary Table S2.

Clinical Features of the Older Brother (III-1)

The age at onset was 17 years, the initial symptom was spastic gait. He gradually developed bradykinesia and postural instability. Levodopa response was positive and aiding with symptoms and the amount of levodopa increased to 500 mg at the age of 28. Wearing off had gradually made his daily life difficult. At the age of

34, he started to take rotigotine. Rotigotine improved his symptoms; however, he started to show impulse control disorder, mainly gambling. As the wearing-off worsened, he underwent a subthalamic nucleus deep brain stimulation (STN-DBS) operation at 37 years old. Although his wearing off had improved, the codes for DBS had to be removed because of a scar infection. Because wearing off was only controlled through medication, DBS surgery was performed again at the age of 39. Since then, DBS and medication have controlled his symptoms well, however, impulse control disorder remained. He did not present cognitive decline throughout the disease's progress. 123-Iodine metaiodobenzylguanidine myocardial scintigraphy showed normal heart-to-mediastinum ratio (early, 2.39, delay, 2.99). Brain magnetic resonance imaging (MRI) was normal, and dopamine transporter single (DAT) photon emission computed tomography (SPECT) showed decreased uptake in both basal ganglia (Fig. 2A). Brain SPECT with N-isopropyl-p[123I]-iodoamphetamine (IMP-SPECT) revealed reduced basal ganglia.

Clinical Features of the Younger Brother (III-2)

The age at onset was 17 years as well for the younger brother. He presented with walking difficulties and right lower limb dystonia. At the first visit to the

neurological clinic at 20, he was considered spastic paraplegia. Levodopa showed significant effects on dystonia. The amount of levodopa increased gradually; he took 500 mg of levodopa at the age of 27. At 31, he presented with diurnal variation of dystonia, bradykinesia, dyskinesia, and posture instability. The Hoehn and Yahr scale was zero at the on phase, three at the off phase. He had started to take levodopa 50 mg frequently (in total, max 500 mg/day), which helped to control his symptoms. Since then, he has been in reasonable control, presenting almost no symptoms at the latest visit at the age of 40. Metaiodobenzylguanidine myocardial scintigraphy showed normal heart-to-mediastinum ratio (early, 2.68, delay, 2.98). Brain MRI was normal, and IMP-SPECT did not show aberrant flow distribution (Fig. 2B).

Exploring the Genome for Potential Causal Variants

DNA was available for the affected twins (III-1, III-2) and mother (II-3), but was not available for the father (II-2). Targeted panel sequencing against known PD-related genes and WES identified a heterozygous *PRKN* variant c.814C > A (p.L272I, rs141366047) in the affected subjects. The allele frequency of c.814C > A is 0.014 by 38KJPN in jMorp,[38] one of the largest genomic databases including the Japanese population, suggesting it is a relatively common single-nucleotide polymorphism in the Japanese population, and therefore, is not likely to be pathogenic. MLPA revealed a heterozygous deletion spanning Exon 3 of *PRKN* in the affected twins, but not in the mother (Fig. 1B). No other variants were identified to be of interest.

Because the affected twins presented typical young onset PD *PRKN* phenotype, we suspected that there might be a complex undetected variant in *PRKN*, which could be missed by short-read WES and MLPA. Therefore, we generated long-read sequencing data for the twins and the mother using Oxford Nanopore Technologies Long-read sequencing. Sample DNA quality control (QC) results were good enough to perform long-read sequencing (Supplementary Fig. S1 and Supplementary Table S3). The overall data output for the samples ranged from 120 to 130 Gb ($\sim$38X coverage assuming a 3.1 Gb human genome), and the average read N50s were 19 kb (Supplementary Table S4). Long-read sequencing confirmed the 60,138 bp deletion (hg38) including exon 3 of *PRKN* (c.(171 + 1_172–1)_ (412 + 1_413-1)del) in the affected twins, but not in the mother, which is concordant with the result of MLPA (Fig. 1C). Additionally long-read sequencing revealed a second mutation, a heterozygotic inversion spanning $\sim$7,425,905 bp (hg38), involving exon 1 to 11 in the affected twins and mother (NC_00006.12: g.161351957_168777862inv). The proximal breakpoint junction was in the intron 11 of *PRKN*, which is predicted to remove exon 12 from the transcript, and
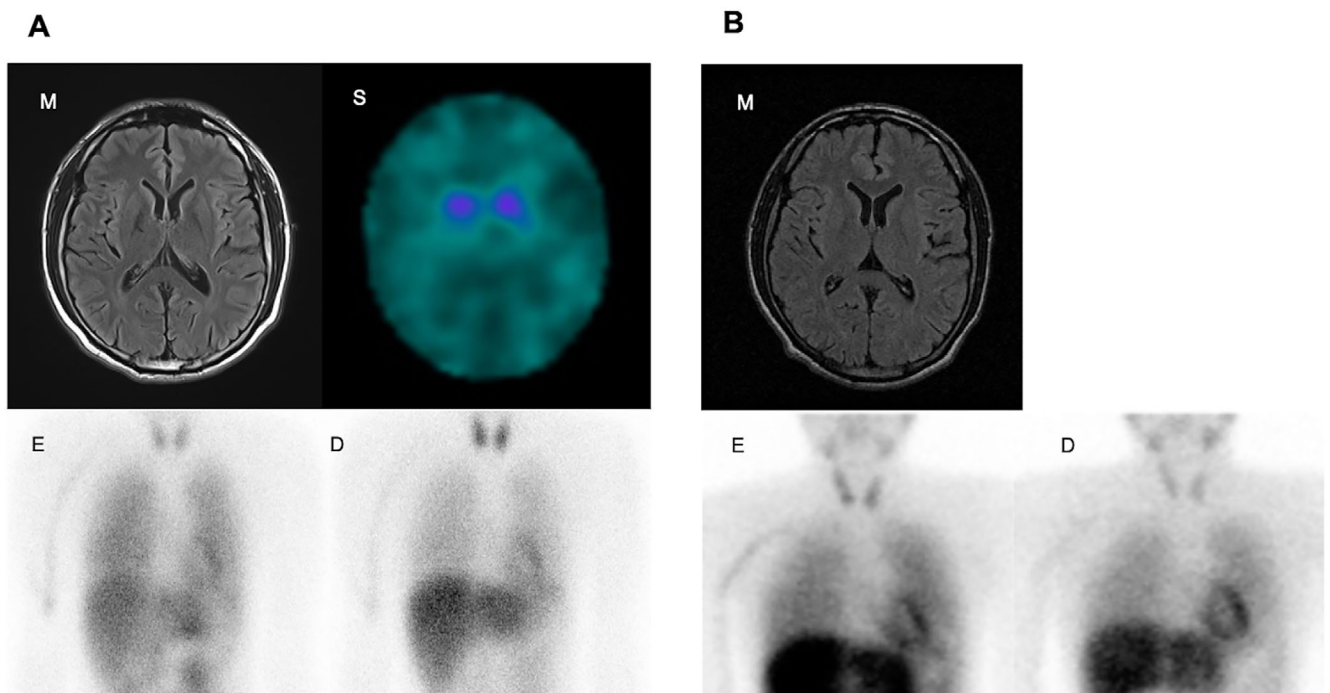

FIG. 2. Clinical Images. (**A**) Images for III-1. (**B**) Images for III-2. M, brain magnetic resonance imaging; S, dopamine transporter single-photon emission computed tomography; E, metaiodobenzylguanidine (MIBG) myocardial scintigraphy on early phase; D, MIBG myocardial scintigraphy on delayed phase.

therefore, resulting in a non-functioning transcript. Importantly, as expected based on MLPA there was no change in the peak of exon 12 in all samples (Fig. 1B). Each of the breakpoints did not have insertion sequences. The allele frequency of this inversion is unknown because no inversion including *PRKN* has been reported in JSV1 from jMorp, a long-read WGS database in the Japanese population.[38] Other than the two *PRKN* variants reported above, we did not identify other potential SVs in PD-related genes. PCR confirmed both breakpoints of the inversion (Supplementary Fig. S2) and additional short-read WGS confirmed the deletion and inversion (Supplementary Figs. S3 and S4).

Breakpoint Region Sequence Analysis of the PRKN Inversion

To consider how the breakpoints occur for this inversion, we performed breakpoint region sequence analysis. Based on the inversion (Fig. 3A), there is a short interspersed nuclear element (SINE) transposable element (148 bp in length) 96 bp upstream of the 5′ breakpoint and a long interspersed nuclear element transposable element (245 bp in length) 178 bp

downstream of the 3′ break point. Sequence similarity was 37.1% in 100 bp up and downstream of the both break points. There was no palindromic sequence around both the breakpoints. Both breakpoints for the deletion of exon 3 (chr6:162,214,329-162,274,466del) were blunt ends. 3′ Breakpoint is in the SINE sequence. Sequence similarity was 42.3% in 100 bp up and downstream of the both break points.

Assessing the Frequency of Large Inversions Involving PRKN Using Short Read WGS from AMP-PD and UK Biobank

Next, we wanted to assess the frequency of large inversions in the *PRKN* genomic region. We used Manta, which is a short read structural variant caller on two large datasets for this: the AMP-PD dataset, which includes 3403 samples (1131 controls and 2272 PD cases) after QC and a subset of the UK Biobank cohort, which has WGS data available on the online research analysis platform including 43,858 participants and including 259 PD cases (based on ICD10 code: G20, PD). In the AMP-PD dataset, we identified 11 inversions in 17 subjects and two of these
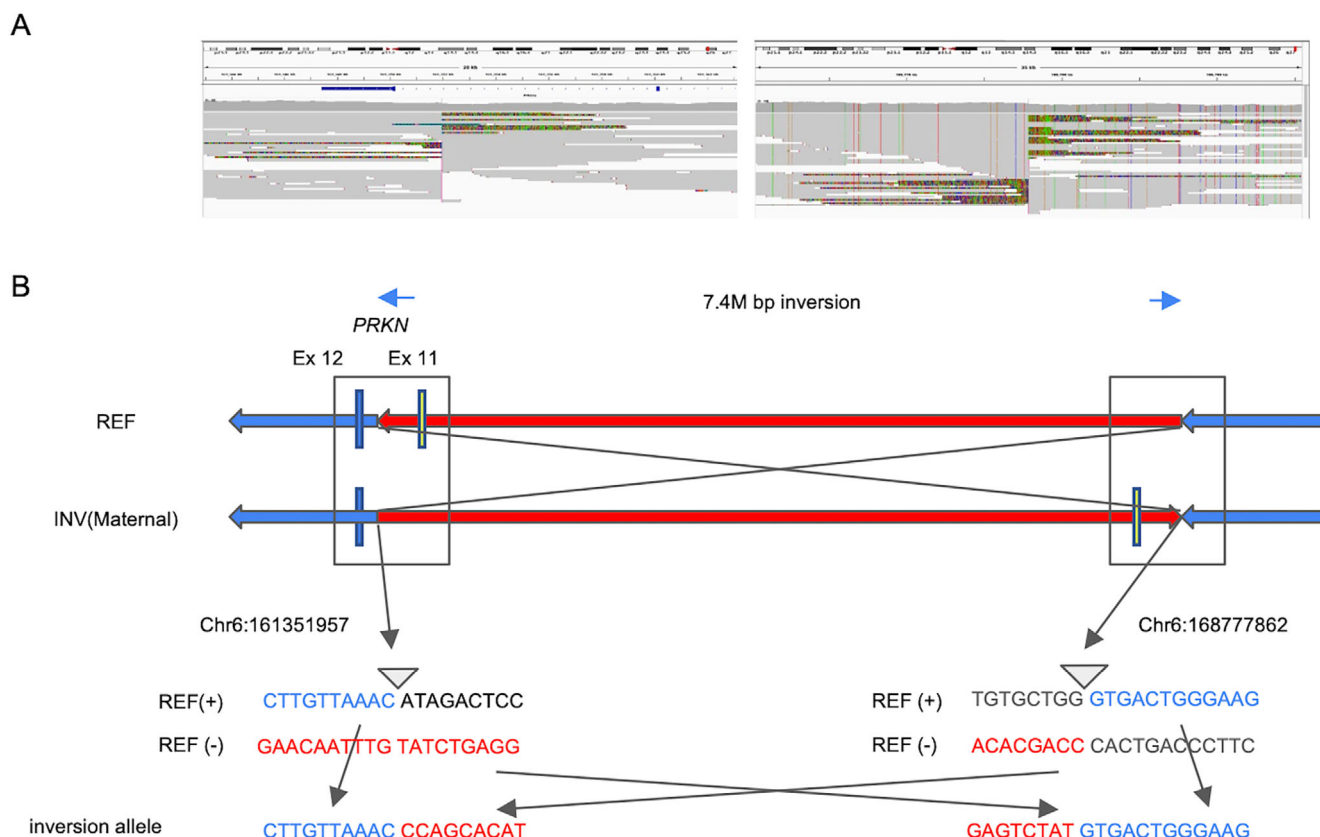

FIG. 3. Detailed description of the identified inversion. (**A**) Screenshot from integrative genomics viewer (IGV) presenting breakpoints of 7 M bp inversion including *Parkin RBR E3 ubiquitin-protein ligase* (*PRKN*). The left figure represents the 5′ break point. The breakpoint is located in intron 11 of *PRKN*. The right figure shows the 3′ breakpoint. (**B**) Schematic representation of the inversion. The upper panel shows the overall picture of the inversion. The lower panel shows the sequence around both breakpoints. 5′ Breakpoint Chr6:161351957, 3′ breakpoint Chr6:168777862. REF, reference; INV, inversion.

11 inversions included a coding exon of PRKN (NM_004562) the others were intronic (Supplementary Table S5). These two inversions affecting exons were called from the same PD subject who was also early onset PD with reported age of onset 42. At the age of 64.5, after 22 years since his onset, he took only 280 mg levodopa per day and his Movement Disorder Society-Sponsored Revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS) scores were not high (MDS-UPDRS I/II/III/IV; 7/1/21/0), showing slow progression of his disease. His Montreal Cognitive Score Assessment score was 29 points suggesting no cognitive decline. Those clinical features were compatible with *PRKN*-PD. However, because this data is based on short-read sequencing we cannot distinguish whether these two inversions are on the same allele or not. We also confirmed he does not have any susceptible pathogenic point mutation in *PRKN* from short-read data.

In the UK Biobank cohort, 43,858 subjects were included and 31 inversions in 46 subjects were found. Nine of the identified inversions included one or more *PRKN* exons and the other 22 were intronic and did not affect a coding exon. Using IGV, one of the nine exon-including inversions was considered to be false-positive because mate reads were mapped to several chromosomes (Supplementary Fig. S5). All the inversions including exon in UK Biobank were identified from non PD subjects. Two inversions in AMP-PD and nine inversions in the UK Biobank were considered potential inversions to affect *PRKN* transcripts.

Discussion

In this study, we identified compound heterozygous SVs in *PRKN* in monozygotic twins with PD. The age of onset of the twins was 17 and the clinical symptoms of the twins were very similar and typical for young onset PD, suggesting a genotype–phenotype correlation. However, only a heterozygotic deletion of exon 3 of *PRKN* was initially identified by short-read sequencing WES and MLPA. Long-read sequencing revealed a large inversion expanding 7.4 M bp with 5′ breakpoint in intron 11 of *PRKN*, removing exon 12 from *PRKN* transcripts. In agreement with recent studies,[20] our study highlights that long-read sequencing can be used to identify new SVs in the PD population, especially in *PRKN*-PD.

The age of onset of the affected twins was 17 years, which was young compared to the median age (31 years) of *PRKN*-PD onset.[17] The motor symptoms were typical for *PRKN*-PD, presenting dyskinesia, motor fluctuations, and a good response to levodopa.[3] Both of the patients did not present dysautonomia and cognitive decline, which is common for *PRKN*-PD.[3] The older brother presented with impulse control disorder, which is one of the psychological symptoms that are not common, as only

2% of the *PRKN*-PD patients presents this.[17] The older brother (III-1) performed STN-DBS operation with an improvement of motor symptoms. This finding was concordant with the literature on DBS suggesting that 76.1% of patients with *PRKN* variants displayed an excellent outcome with DBS.[39]

To date, over a hundred of *PRKN* causative variants (36 deletions, 22 duplications, 81 SNVs or indels, https://www.mdsgene.org/, February 2023) are reported and only one inversion is reported including exon 2 to exon 5, which is different from the one identified in our study (c.8-224652_618 + 32307dupinvAAGATTTins). This inversion was identified from young onset PD patients in Poland.[32] When searching Clinvar (https://www.ncbi.nlm.nih.gov/clinvar/?term=PRKN%5Bsym%5D), one inversion (c.618 + 7842_618 + 7934inv) in intron 5 was registered in *PRKN*. There was another case report describing this inversion in *PRKN*. A report from Israel describes the pathogenic homozygous inversion of exon 5 in *PRKN*.[40] They report a consanguineous family with four of 12 siblings developed young onset dystonia-parkinsonism. The symptoms were severe, with a median age at onset of 11 years. Dystonia generalized gradually to the other limbs. A few years after the onset, they presented parkinsonism with motor fluctuation. Three of four affected individuals had subthalamic DBS. Two of the affected subjects had 18F-l-3,4-dihydroxyphenylalanine brain positron emission tomography/computed tomography studies and showed impaired striatal presynaptic dopaminergic function. Using short read WGS, the affected subjects were shown to harbor biallelic inversion including exon 5 followed by 49 kb deletion, which causes skipping of exon 5 in the cDNA.

Although the role of biallelic variants in *PRKN* is well established, the effects of heterozygous *PRKN* in PD are conflicting.[11-16] Meta-analysis of *PRKN* heterozygous suggests an association between the heterozygous *PRKN* variant and PD. However, the potential confounding effect of an unrevealed second mutation is pointed out.[14] Recently another study showed no association between heterozygous *PRKN* and PD by comparing short-read sequencing WGS.[16] One potential explanation is that there could be other complex SVs that affect *PRKN* function or deep intronic variants, which affect splicing that are missed in previous genotyping technologies. We speculate that some PD cases harboring heterozygous *PRKN* variants may have a complex inversion as a second variant, especially in the young-onset PD following an autosomal recessive pattern. As our cases had been classified as heterozygous *PRKN* carriers until the large inversion was identified, there are likely other young onset PD cases carrying unrevealed SVs, especially inversions or other complex SVs. Our previous report shows 2.5% of familial PD cases harbor *PRKN* heterozygous variants.[41]

Next, we surveyed large scale short-read WGS datasets to screen similar inversions in the general population

including also PD cohorts. Overall, inversions of *PRKN* in both the UK Biobank and AMP-PD datasets were rare, and also we did not identify the exact inversion or a similar size inversion as the reported *PRKN* family. However, all identified *PRKN* inversions are likely pathogenic and, therefore, of interest. It is interesting that the age of PD onset is relatively young (42 years) in the subject who harbored two inversions in AMP-PD dataset. Naturally, the low frequency of *PRKN* inversions might be due to inversions itself being quite rare in *PRKN*, but it is also likely that short-read sequencing could not identify all SVs. For example, it is known that repeats or transposable elements can cause SVs, and short read sequencing methods often struggle to map accurately in those regions.[42,43]

As with any study, there are some limitations in the presented work. First, we did not have access to RNA sequencing data to assess the effect of the inversion on the transcript levels and isoforms. However, given the predicted structure of the inversion, it is likely that all *PRKN* isoforms including exon 12 will result in a nonfunctioning transcript given the lack of a stop codon. We are planning to investigate this effect using differentiated neurons derived from induced pluripotent stem cell. Second, the DNA of the father (II-2) was not available. However, as the non-affected mother (II-3) only harbors the large inversion, including *PRKN*, it is very likely that the father is carrier of the exon 3 deletion. Third, in our assessment of the frequency of potential damaging *PRKN* inversion in the general population we only included European ancestry individuals. Future studies should explore the frequency of *PRKN* inversions across populations and especially in unresolved early onset PD cases.

In summary, we report how long-read sequencing can identify complex *PRKN* SVs, which are likely to be missed by MLPA and conventional short-read sequencing methods. We expect that several other early-onset PD cases with a *PRKN* phenotype have a second complex variant that long-read sequencing can resolve. This study emphasizes the usefulness of long-read sequencing in the research of familial PD cases. ∎

Data Availability Statement

UK Biobank data is available upon application at the UK Biobank website (https://www.ukbiobank.ac.uk/). AMP-PD data is available upon application at the AMP-PD website (https://www.amp-pd.org/).

References

1. Poewe W, Seppi K, Tanner CM, et al. Parkinson disease. Nat Rev Dis Primers 2017;3:17013.

2. Mehanna R, Smilowska K, Fleisher J, et al. Age cutoff for early-onset Parkinson's disease: recommendations from the International Parkinson and Movement Disorder Society task force on early onset Parkinson's disease. Mov Disord Clin Pract 2022;9(7):869–878.

3. Lesage S, Lunati A, Houot M, et al. Characterization of recessive Parkinson disease in a large multicenter study. Ann Neurol 2020;88(4):843–850.

4. Kitada T, Asakawa S, Hattori N, et al. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. Nature 1998;392(6676):605–608. https://doi.org/10.1038/33416

5. Matsuda N, Sato S, Shiba K, et al. PINK1 stabilized by mitochondrial depolarization recruits parkin to damaged mitochondria and activates latent parkin for mitophagy. J Cell Biol 2010;189(2):211–221.

6. Smith DI, Zhu Y, McAvoy S, Kuhn R. Common fragile sites, extremely large genes, neural development and cancer. Cancer Lett 2006;232(1):48–57.

7. Denison SR, Callahan G, Becker NA, Phillips LA, Smith DI. Characterization of FRA6E and its potential role in autosomal recessive juvenile parkinsonism and ovarian cancer. Genes Chromosomes Cancer 2003;38(1):40–52.

8. Durkin SG, Glover TW. Chromosome fragile sites. Annu Rev Genet 2007;41:169–192.

9. Kilarski LL, Pearson JP, Newsway V, et al. Systematic review and UK-based study of PARK2 (parkin), PINK1, PARK7 (DJ-1) and LRRK2 in early-onset Parkinson's disease. Mov Disord 2012;27(12):1522–1529.

10. Zhao Y, Qin L, Pan H, et al. The role of genetics in Parkinson's disease: a large cohort study in Chinese mainland population. Brain 2020;143(7):2220–2234. https://doi.org/10.1093/brain/awaa167

11. Yu E, Rudakou U, Krohn L, et al. Analysis of heterozygous PRKN variants and copy-number variations in Parkinson's disease. Mov Disord 2021;36(1):178–187.

12. Kay DM, Stevens CF, Hamza TH, et al. A comprehensive analysis of deletions, multiplications, and copy number variations in PARK2. Neurology 2010;75(13):1189–1194.

13. Lesage S, Lohmann E, Tison F, et al. Rare heterozygous parkin variants in French early-onset Parkinson disease patients and controls. J Med Genet 2008;45(1):43–46.

14. Lubbe SJ, Bustos BI, Hu J, et al. Assessing the relationship between monoallelic PRKN mutations and Parkinson's risk. Hum Mol Genet 2021;30(1):78–86.

15. Huttenlocher J, Stefansson H, Steinberg S, et al. Heterozygote carriers for CNVs in PARK2 are at increased risk of Parkinson's disease. Hum Mol Genet 2015;24(19):5637–5643.

16. Zhu W, Huang X, Yoon E, et al. Heterozygous PRKN mutations are common but do not increase the risk of Parkinson's disease. Brain 2022;145(6):2077–2091.

17. Kasten M, Hartmann C, Hampf J, et al. Genotype-phenotype relations for the Parkinson's disease genes parkin, PINK1, DJ1: MDSGene systematic review. Mov Disord 2018;33(5):730–741.

18. Schneider SA, Alcalay RN. Neuropathology of genetic synucleinopathies with parkinsonism: review of the literature. Mov Disord 2017;32(11):1504–1523.

19. Mantere T, Kersten S, Hoischen A. Long-read sequencing emerging in medical genetics. Front Genet 2019;10:426.

20. Billingsley KJ, Ding J, Jerez PA, et al. Genome-wide analysis of structural variants in Parkinson disease. Ann Neurol 2023;93:1012–1022. https://doi.org/10.1002/ana.26608

21. Gibb WR, Lees AJ. The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease. J Neurol Neurosurg Psychiatry 1988;51(6):745–752. https://doi.org/10.1136/jnnp.51.6.745

22. Postuma RB, Berg D, Stern M, et al. MDS clinical diagnostic criteria for Parkinson's disease. Mov Disord 2015;30(12):1591–1601.

23. Daida K, Nishioka K, Li Y, et al. PLA2G6 variants associated with the number of affected alleles in Parkinson's disease in Japan. Neurobiol Aging 2021;97:147.e1–e147.e9.

24. Van der Auwera GA, O'Connor BD. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra. California: O'Reilly Media; 2020.

25. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38(16):e164.

26. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 2020;581(7809):434–443.

27. Billingsley KJ, Jerez PA, Miano-Burkhardt A, et al. Processing frozen human blood samples for population-scale Oxford nanopore long-read DNA sequencing SOP. protocols.io 2022. https://dx.doi.org/10.17504/protocols.io.ewov1n93ygr2/v1

28. Wick RR, Judd LM, Holt KE. Performance of neural network basecalling tools for Oxford nanopore sequencing. Genome Biol 2019;20(1):129.

29. Jain C, Rhie A, Hansen NF, Koren S, Phillippy AM. Long-read mapping to repetitive reference sequences using Winnowmap2. Nat Methods 2022;19(6):705–710.

30. Sedlazeck FJ, Rescheneder P, Smolka M, et al. Accurate detection of complex structural variations using single-molecule sequencing. Nat Methods 2018;15(6):461–468.

31. Geoffroy V, Herenger Y, Kress A, et al. AnnotSV: an integrated tool for structural variations annotation. Bioinformatics 2018;34(20):3572–3574.

32. Ambroziak W, Koziorowski D, Duszyc K, et al. Genomic instability in the PARK2 locus is associated with Parkinson's disease. J Appl Genet 2015;56(4):451–461.

33. Mitsui J, Takahashi Y, Goto J, et al. Mechanisms of genomic instabilities underlying two common fragile-site-associated loci, PARK2 and DMD, in germ cell and cancer cell lines. Am J Hum Genet 2010;87(1):75–89.

34. Madeira F, Pearce M, Tivey ARN, et al. Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic Acids Res 2022;50(W1):W276–W279.

35. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics 2009;25(14):1754–1760.

36. Chen X, Schulz-Trieglaff O, Shaw R, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics 2016;32(8):1220–1222.

37. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 2013;14(2):178–192.

38. Tadaka S, Hishinuma E, Komaki S, et al. jMorp updates in 2020: large enhancement of multi-omics data resources on the general Japanese population. Nucleic Acids Res 2021;49(D1):D536–D544.

39. Kuusimäki T, Korpela J, Pekkonen E, Martikainen MH, Antonini A, Kaasinen V. Deep brain stimulation for monogenic Parkinson's disease: a systematic review. J Neurol 2020;267(4):883–897.

40. Mor-Shaked H, Paz-Ebstein E, Basal A, et al. Levodopa-responsive dystonia caused by biallelic exon inversion invisible to exome sequencing. Brain Commun 2021;3(3):fcab197.

41. Yoshino H, Li Y, Nishioka K, et al. Genotype-phenotype correlation of Parkinson's disease with PRKN variants. Neurobiol Aging 2022;114:117–128.

42. Payer LM, Burns KH. Transposable elements in human genetic disease. Nat Rev Genet 2019;20(12):760–772.

43. Mahmoud M, Gobet N, Cruz-Dávalos DI, Mounier N, Dessimoz C, Sedlazeck FJ. Structural variant calling: the long and the short of it. Genome Biol 2019;20(1):246.

Supporting Data

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.