HHS Public Access

Author manuscript

Mov Disord. Author manuscript; available in PMC 2020 July 14.

Published in final edited form as:

Mov Disord. 2020 May ; 35(5): 868–876. doi:10.1002/mds.27998.

Replication-based rearrangements are a major mechanism for *SNCA* duplication in Parkinson disease

Soo Hyun Seo, MD[1], **Albino Bacolla, PhD**[2], **Dallah Yoo, MD**[3], **Yoon Jung Koo, BS**[4], **Sung Im Cho, MS**[4], **Man Jin Kim, MD**[4], **Moon-Woo Seong, MD, PhD**[4], **Han-Joon Kim, MD, PhD**[5], **Jong-Min Kim, MD, PhD**[6], **John A. Tainer, PhD**[2], **Sung Sup Park, MD, PhD**[4,7], **Ji Yeon Kim, MD, PhD**[7,*], **Beomseok. Jeon, MD, PhD**[5]

[1]Department of Laboratory Medicine, Seoul National University Bundang Hospital, Seongnam, Korea; Seoul National University College of Medicine, Seoul, Korea

[2]Department of Molecular and Cellular Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

[3]Department of Neurology, Kyung Hee University Hospital, Seoul, Korea

[4]Department of Laboratory Medicine, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Korea

[5]Department of Neurology, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Korea

[6]Department of Neurology, Seoul National University Bundang Hospital, Seongnam, Korea; Seoul National University College of Medicine, Seoul, Korea

[7]Biomedical Research Institute, Seoul National University Hospital, Seoul, Korea

Abstract

Background—*SNCA* multiplication is a genomic cause of familial Parkinson disease, showing dosage-dependent toxicity. Until now, non-allelic homologous recombination was suggested as the

*Corresponding Authors: Ji Yeon Kim, Biomedical Research Institute, Seoul National University Hospital, Seoul, Korea, jkimmd@snu.ac.kr.

Authors' Roles:
1) Research project: A. Conception, B. Organization, C. Execution;
2) Statistical Analysis: A. Design, B. Execution, C. Review and Critique;
3) Manuscript: A. Writing of the first draft, B. Review and Critique.
S.H.S: 1A, 2B, 2A, 2B, 3A
A.B: 2C, 3B
D.Y: 1C, 2C
Y.J.K: 1C
S.I.C: 1B, 1C
M.J.K: 2C
M.-W.S: 1B, 2C
H.-J.K: 2C
J.-M.K: 2C
J.A.T: 2C, 3B
S.S.P: 1A, 2C, 3B
J.Y.K: 1A, 2C, 3B
B.S.J: 1A, 2C, 3B

Financial disclosures/Conflict of interest: Nothing to report.

mechanism of *SNCA* duplication, based on various types of repetitive elements found in the spanning region of the breakpoints. However, the sequence at the breakpoint was analyzed only for one case.

Objectives—We have analyzed the breakpoint sequences of 6 patients with Parkinson disease who had duplicated *SNCA,* using whole genome sequencing data to elucidate the mechanism of *SNCA* duplication.

Methods—Six patient samples with *SNCA* duplication underwent whole genome sequencing. The duplicated regions were defined with nucleotide-resolution breakpoints, which were confirmed by junction PCR and Sanger sequencing. The search for potential non-B DNA-forming sequences and stem-loop structure predictions were conducted.

Results—Duplicated regions ranged from the smallest region of 718.3 kb to the largest one of 4,162 kb. Repetitive elements were found at 8 of the 12 breakpoint sequences on each side of the junction, but none of the pairs shared overt homologies. Five of these 6 junctions had microhomologies (2–4 bp) at the breakpoint, and a short stretch of sequences were inserted in 3 cases. All except one junction were located within or next to stem-loop structures.

Conclusion—Our study has determined that homologous recombination mechanisms involving repetitive elements are not the main cause for the duplication of *SNCA*. The presence of microhomology at the junctions and their position within stem-loop structures suggest that FoSTeS/MMBIR may be a common mechanism for *SNCA* amplification.

Keywords

SNCA duplication; Parkinson disease; FoSTeS; MMBIR; microhomology

Introduction

Parkinson disease (PD) is the second most common neurodegenerative disorder and is characterized by the clinical motor symptoms of tremor, rigidity, and bradykinesia[1, 2]. Non-motor symptoms such as insomnia, depression, rapid eye movement sleep behavior disorder and dysautonomia are commonly associated with more advanced stages of the disease. As the disease progresses, cognitive impairments or visual hallucinations may occur. Only about 10% of PD cases are known to be linked to a genetic cause and several genes, such as *LRRK2, SNCA, PRKN, PINK1,* and *PARK7*, have been identified as the cause of hereditary PD[1].

SNCA is one of the genes responsible for autosomal dominant PD. Missense variants in *SNCA* were the first genetic causes identified as the hereditary form of the disease[3, 4]. However, whole gene multiplications had been more frequently detected than the single nucleotide variants (SNV). It has been reported that multiplications of *SNCA* are observed in around 0.05% of European PD patient populations.[5, 6] Other genes related to hereditary forms of PD have copy number variants (CNVs) demonstrating the role of loss of function.[7] CNVs in *SNCA,* however, seem to suggest a gain of function. Interestingly, CNVs in *SNCA* have been limited to multiplication of the whole gene (with no small sized CNVs) and they are known to increase mRNA expression [8], showing dosage-dependent toxicity.[9]

The formation of structural variants may arise from different mutational mechanism[10]. Among the recombination-based mechanisms, the prevailing mechanism for recurrent and some nonrecurrent rearrangements is nonallelic homologous recombination (NAHR), and other nonrecurrent rearrangements mechanisms are mediated by non-homologous end joining (NHEJ). NAHR uses repetitive sequences or low-copy repeats (LCRs) as homologous recombination substrates whereas NHEJ usually generates simple, blunt end points or a short stretch of microhomology. Other rearrangement mechanisms requiring microhomology are also possible for nonrecurrent recombination, such as fork stalling and template switching (FoSTeS)[11] or microhomology-medicated break-induced replication (MMBIR)[12]. These replication-based mechanisms result from template switching while replication fork is stalled.

Until now, non-allelic homologous recombination (NAHR) was suggested as the mechanism of *SNCA* multiplications based on various types of repetitive elements found in the spanning region of breakpoints[13, 14]. Yet, most of the cases did not have their exact breakpoint sequence mapped and only crude approximate boundaries were determined by low resolution array CGH methods. Only one case had breakpoint analysis at a base-pair level, showing the homology between repetitive elements on both sides of the junction.[15] Since the underlying rearrangement mechanism of *SNCA* duplication remains yet to be explored, we have analyzed the genomic data of six PD patients with duplicated *SNCA*. We have defined the exact range of the duplicated regions, analyzed the junction sequences, and described the clinical characteristics of each patient.

Methods

Parkinson disease patients tested for *SNCA* duplications

A total of 408 patients diagnosed with PD were included in the study. None of them were previously tested for *SNCA* multiplication. DNA was extracted from whole blood collected in EDTA-containing tubes. Additionally, five samples confirmed to have duplication from previous studies[16, 17] were included for whole genome sequencing. Informed consent was obtained from all participants and all experiments were performed according to the Declaration of Helsinki. The study was approved by the institutional review board of Seoul National University Hospital (IRB- H-1805–157-948).

Primary detection of *SNCA* duplications

Gene dosage of *SNCA* was assessed by semiquantitative multiplex PCR as previously described.[16] To confirm the *SNCA* duplications, multiplex ligation-dependent probe amplification (MLPA) was performed using the SALSA MLPA P051 Parkinson mix 1 probemix (MRC-Holland, Amsterdam, The Netherlands). PCR products were analyzed on an ABI 3130 analyzer with Genemarker ver. 1.51 (Softgenetics, State College, PA, USA).

Whole Genome Sequencing (WGS) and breakpoint determination

Six samples confirmed with *SNCA* duplication underwent whole genome sequencing. Paired-end libraries of average read-length 149 bp were prepared using the TruSeq Nano DNA library generation protocol and sequenced on the HiSeq X Ten platform (Illumina Inc.,

San Diego, CA, USA). WGS data were aligned using Isaac Aligner[18] and Control-FREEC was used to calculate the copy number and allelic content[19]. Variant calling was obtained using NextGENe software (SoftGenetics, State College, PA, USA). Duplicated regions containing the *SCNA* were visualized using Integrative Genomics Viewer (IGV) by generating a tiled data file (TDF). Discordant read pairs were taken into consideration and evaluation of the exact breakpoint of the duplication was performed by manually analyzing the read sequences of split reads. To confirm the breakpoint sequences, junction PCR and Sanger sequencing were performed. Specific primers were designed for each junction and sequencing was performed using the ABI 3730 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Among the polymorphisms located within and flanking the *SNCA* gene, those known to be associated with PD risk[20, 21] (rs181489, rs356219, rs11931074, rs356220, rs356165, rs2736990, rs356186, rs2737029, rs894278, rs10005233, rs2619364, rs2583988) were assessed from WGS data for each case and were used to define a haplotype as well.

Junction sequence analyses

The dataset of human SNPs was from the NCBI dbSNP Build 151, which was downloaded from UCSC at http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/ (file snp151.txt.gz). We selected the SWEGEN-tagged “genomic single” SNP entries, which contained 29,691,973 genetic variants recorded in the Swedish population[22], 1,996,894 of which occurred on chromosome 4.

Repetitive elements in the breakpoint regions were searched using RepeatMasker from the UCSC genome browser. Sequence similarities were compared using the BLAST program (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

The search for potential non-B DNA-forming sequences (PONDS) included direct repeats (looped-out structures), inverted repeats (hairpins), four or more runs of GGG each separated by 1–7 bases (quadruplex DNA), purine•pyrimidine tracts with mirror repeat symmetry (triplex DNA) and alternating purine-pyrimidine tracts (left-handed Z-DNA). The search was performed with in-house scripts[23] on 1-kb sequences in hg19 coordinates, each containing the duplication junctions at the center. The same search was conducted on additional 20,282 1-kb sequences chosen at random, which served as control. Occasional controls with sequence gaps were excluded. To assess the statistical significance between cases and controls, we computed the average number of PONDS per 100 bp in the 10 junction fragments from the cases and ~20,200 controls and performed z-tests.

Stem-loop structure predictions were conducted on 201-bp sequences with the junction in the middle using Mfold (http://unafold.rna.albany.edu/?q=mfold). These sequence intervals were also subjected to intrinsic DNA curvature analysis[24, 25] using https://www.lfd.uci.edu/~gohlke/dnacurve/ and EMBOSS fuzznuc (http://www.bioinformatics.nl/cgi-bin/emboss/fuzznuc); PDB rendering was attained with PyMOL (https://pymol.org/2/). Dot plots for the visualization of sequence matrix identities between proximal and distal junctions were obtained with EMBOSS dotmatcher (http://www.bioinformatics.nl/cgi-bin/emboss/dotmatcher) on the 1-kb sequences.

Results

SNCA duplication detected by semiquantitative multiplex PCR and MLPA

Among 408 patient samples tested for *SNCA* multiplication, only one sample had triple copies of *SNCA* interpreted as a duplication of the gene. Together with two previous studies that reported duplications of *SNCA* in patients with Parkinson disease from the same single center as in this study, a total of 2031 patients were screened for *SNCA* CNVs. Among them, a total of 6 were found to have a duplication. This results in a frequency of 0.30% of the whole cohort of PD.

Characterization of duplicated regions

WGS data of six samples with *SNCA* duplication showed mean depth coverage of 35.4X across the whole genome, and 53.5X in the duplicated regions in each patient. 99.2% of the duplicated regions was covered at ≥20X. Duplicated regions of each sample encompassing *SNCA* were shown in IGV (Figure 1). All duplications were in tandem and in direct orientation, and the critical region encompassed only *SNCA*. The smallest size of the duplication region was 718.3 kb, detected in Case 2, which included 3 genes, *SNCA, MMRN1* and *CCSER1*. The largest was 4,162 kb, detected in Case 4, and included 10 genes, *HERC3*, *NAP1L5*, *FAM13A*, *TIGD2*, *GPRIN3*, *SNCA*, *MMRN1*, *CCSER1*, *LNCPRESS2,* and *GRID2*. (Table 1). In the breakpoint region of Case 4, *GRID2* and *HERC3* were found to be localized at the breakpoint junction, which are expected to form a fused protein with intron 2 of *GRID2* joined to intron 12 of *HERC3*.

Identification of duplication junction sequences

Junction sequences were identified by PCR followed by Sanger sequencing using the primers specifically designed for each case (Supplementary Table 1). Healthy control samples were used to confirm that amplification only occurred in the specific samples (Supplementary Figure 1). Junction PCR and sequencing confirmed the exact sequence of the breakpoints (Figure 2).

Eight of the 12 analyzed breakpoint sequences on each side of the junction mapped within a specific repetitive element (Supplementary Table 2, Supplementary Figure 2), and included six long interspersed elements (LINEs), one short interspersed element (SINE) and one long terminal repeat (LTR). Case 2 and 5 had repetitive elements on both sides of the breakpoint (Figure 2), but there were no overt homologies between the sequences. No Alu elements were found in any of the regions. None of the cases shared a significant homology between the sequences within a range of 1 kb around the breakpoints.

Five of these six junctions had microhomologies at the breakpoint, ranging from 2 to 4 nucleotides. Three of the junctions had a short stretch of sequences inserted (1 to 23 bp). Among them, two of the inserted sequences (Case 1 and 4) were homologous to the sequence right next to the breakpoint (Figure 2). The other inserted sequence observed in Case 5 showed a stretch of 23 bps that was not homologous to the sequences near the junction, but it contained some repeated TGs, which could be from a simple repeat sequence of (TG)n located 152 bp downstream of the distal breakpoint or from other genomic regions.

Analyses for potential stem-loop structures prediction

Local rates of hypermutation and DNA structural features have been shown to predispose to non-recurrent structural variations[26–30]. A map of common SNPs along chromosome 4 (Figure 3) revealed three hypermutability regions, two at the subdistal ends and a third near the periproximal p-arm (Figure 3A), thereby excluding a role for SNP hypermutation as a cause for *SNCA* amplification. Next, given that NAHR between LINE-1 elements was reported to have mediated a case of *SNCA* amplification[15], we assessed the extent of sequence identity within ±500 bp of the duplication junctions between the proximal and distal sequences. Despite the consistent presence of repetitive elements (Supplementary Figure 2), no significant homologies were discernable (Supplementary Figure 3). Next, we examined the occurrence of static bending near (within ±100 bp) the duplication junctions, which would suggest local hypermutability and increased DNA melting[23]. DNA bending was prominent around two junctions in cases 3/6 and 4 (Supplementary Figure 3 A–C), but was modest around the other 8 junctions (Supplementary Figure 3 D–F). Finally, we computed the density of motifs known to form potential non-B DNA structures (PONDS)[23, 31] within ±100 bp of the junctions. No enrichment in the 5 types of PONDS was observed (Figure 3B), although a short inverted repeat (TGGGTAT/ATACCCA) and two triplex-forming repeats (AGAGGA(N7)AGGAGA and AAGAAA(N5)AAAGAA) in Cases 3/6 and a triplex-forming repeat (A(6)N(3)(A(6)) in Case 4 were within ~4 helical turns of the junctions. By contrast, a search for the potential folding of imperfect inverted repeats into stem-loop structures revealed that, with one exception, all junctions were located within or next to stem-loop structures. Therefore, we conclude that small imperfect cruciforms may have potentiated the formation of *SNCA* duplications.

Haplotype analysis of duplicated regions

Without overt homologies observed at the junction, duplications involving *SNCA* were expected to be non-recurrent episodes; yet Case 3 and 6 shared the exact same breakpoint sequence. These two probands were not related, so haplotype was assessed to check for a possible founder effect. Twelve SNPs encompassing the *SNCA* gene were probed using IGV and a combination of each allele with higher number of reads was considered as one haplotype of duplicated alleles, while lower read alleles were considered as another (Supplementary Table 3). A haplotype with higher number of reads in Case 3 was different from that of Case 6 by two SNPs, suggesting that a common founder between the two cases seems unlikely.

Variants in other related genes

Aside from the duplications of *SNCA*, other single nucleotide variants (SNVs) and small indel variants of the other known PD-related genes were examined using the WGS data. A total of 38 genes (*ATP13A2, ATP1A3, ATXN2, C19orf12, CHCHD2, CP, DNAJC6, DNAJC13, FA2H, FBXO7, FTL, GBA, GCH1, GLB1, HTRA2, LRRK2, PANK2, PARK7, PINK1, PLA2G6, POLG, PRKN, PTS, QDPR, RAB39B, SLC30A10, SLC6A3, SNCA, SPR, SYNJ1, TAF1, TH, TMEM230, UCHL1, VPS13C, VPS35, WDR45, ZFYVE26)* related to several conditions of parkinsonism[32] were searched for additional pathogenic variants. No pathogenic variants were detected in the coding exons with 10 bp of intronic

flanking regions, while Case 4 carried an intron variant of unknown significance (VUS) in *ATP13A2* (c.2529+9G>A).

Clinical characteristics

Clinical characteristics of the 6 cases of duplication were examined thoroughly (Table 2). Cognitive impairment was found in all cases except in Case 2. Visual hallucination developed in four of the patients (Cases 1, 4, 5 and 6). When considering the risk alleles of 12 SNPs, Case 1 who carried most of the risk alleles had earliest onset (at the age of 40), while Case 3 with the least number of risk alleles had the latest onset of PD (at the age of 66). The UPDRS motor score was available for 4 patients and the Hoehn and Yahr scale was available for all patients. Case 4 showed the highest UPDRS score of 36, but also had the longest disease duration of 5 years. Disease severity showed no correlation with *SNCA* duplication size.

Discussion

The 4q22 region has been regarded as inherently prone to disruption due to various repetitive elements and homologies, which were predicted to promote NAHR. The first case of *SNCA* duplication that had an exact breakpoint mapped showed >95% homology between LINEs residing at both breakpoints, suggesting that NAHR was the underlying mechanism[15]. Meanwhile, other studies had not revealed the exact sequence of the breakpoint junctions and only suggested the role of repetitive elements located in the window regions narrowed by low resolution array CGH[13, 14]. The result of our study has shown that there were no significant homologies between proximal and distal breakpoints and microhomologies were frequently observed at the duplication junctions, along with sequence insertions. The presence of microhomology at the junctions and their position within stem-loop structures supports the concept that *SNCA* duplication arose from stalled replication forks, which then restarted at ectopic sites upon minimal pairing with an acceptor template, as described by the FoSTeS/MMBIR models of replication repair[11, 12, 26]. Template switching may be reiterative, as suggested by Case 5, where the micro-insertion may have been acquired from either chromosome 6 or 21, and Case 4, in which the ATTTGTCAA motif was copied twice. Stem-loop structures may oppose strand reannealing during transcription and replication, thereby proving an opportunity for DNA damage at unpaired bases or cleavage by structure-specific nucleases, lesions which may halt replication. Alternatively, stem-loops may provide hubs for the re-priming of forks that have stalled at random damaged sites. This conclusion is similar to what was observed in the Iowa kindred case of *SNCA* triplication described by Zafar et al[33], in which there was a small region of a duplication detected on both side of the triplicated region, suggesting two independent mutational events. It is possible that this complex structural variant also resulted from MMBIR[10].

It is interesting to note that the proximal junction of recurrent Cases 3 and 6 was also flanked by triplex-forming motifs, whereas at the distal junction a perfect inverted repeat overlapped with a larger stem-loop. In this context it will be therefore relevant to examine larger metadata to assess the extent of DNA structure complexity that would be required to elicit recurrent rearrangements.

Compared to the previously reported proportion of *SNCA* duplications occurring in 0.05% of the general PD cohort, our data showed 0.30% in a cohort of a single medical center, which was considerably higher than expected. This could possibly be due to founder effects, but no evidence was detected to support this hypothesis. Looking into the junction region, Case 4 had both breakpoints each disrupting *HERC3* and *GRID2*. We predict that intron 2 of *GRID2* would join intron 12 of *HERC3* to generate a fusion protein. *GRID2* had been previously reported as a genetic cause for a neurodegenerative disease, and SNVs as well as gross deletions in *GRID2* have been reported in cerebellar ataxia.[34] However, whether this fusion gene created at the duplication junction site will affect the function remains unclear.

PD patients with *SNCA* multiplication are known for early onset age and prominent dementia[32], compared to the classical parkinsonism cases with *SNCA* missense variants. Also, symptoms are known to be sensitive to dosage-effects arising from *SNCA* multiplication[9, 35]. Patients with *SNCA* triplication commonly develop dementia, while the frequency is slightly lower in cases with duplication. Most of the patients included in this study developed dementia at some point during disease progression. Since the onset age of dementia or cognitive dysfunction in duplication cases are not as early as those with triplication, it may appear as lower in frequency when addressed at the early course of the disease. However, when followed through the disease course they are likely to eventually progress towards a cognitive decline. In this study, Case 5 had the most severe cognitive deficit (MMSE 8 at the time of last follow up). Case 2, who carried the smallest duplicated region, showed relatively mild progression over 15 years of follow up, with no cognitive decline, no visual hallucination nor psychosis. He showed a UPDRS motor score of 31 when assessed 2 years after the onset. Yet, no significant differences in clinical characteristics that can be related to the genomic features were observed.

Among the patients in this study, only Case 1 had a family history of parkinsonism, his mother. Case 2 and 3 had some of their family members tested for *SNCA* duplication, and an older brother of Case 2 and a daughter of Case 3 were also confirmed with the duplication. The older brother of Case 2 had been examined thoroughly, but it was concluded that he had ataxia rather than parkinsonism. The daughter of Case 3 did not have any symptoms of parkinsonism at the age of 44, but she was not followed up afterward. This is consistent with the previous finding that the penetrance of the *SNCA* duplication is incomplete and is likely age-dependent.[9, 36]

This study has the limitations of being a single-center study, which is restricted to a geographical and ethnic cohort, and of not including cases of *SNCA* triplication. Future studies in other cohorts will be necessary for assess whether other mechanisms are involved in *SNCA* duplications.

In conclusion, we carried out a breakpoint analysis of the junctions in six Parkinson disease patients with *SNCA* duplication. We have identified the junction sequences at the base-pair level, and in contrast to previous reports, we showed that NAHR was not the major mechanism involved in the duplication of *SNCA*. Despite the presence of repetitive elements, we did not find significant homologies between proximal and distal breakpoints. Instead, we found the presence of microhomology at the junctions and their position within

stem-loop structures, suggesting that FoSTeS/MMBIR may be a common mechanism for SNCA amplification.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

This research was supported in part by the National Cancer Institute grant CA092584 to J.A.T. This work used the Extreme Science and Engineering Discovery Environment (XSEDE) Bridges at the Pittsburgh Supercomputing Center through allocation MCB170053. We thank Dr. Davide Moiani for help with PyMOL.

Funding sources: Nothing to report.

Financial Disclosures of all authors (for the preceding 12 months):

Dr. Beom Seok Jeon has received research support Seoul National University College of Medicine, Seoul National University Hospital, Sinyang Cultural Foundation, Peptron, and Abbvie Korea.

Dr. Sung Sup Park has received research support from Korea Rare-disease Genetic Diagnostics Project (KRGDP), Korea Centers for Disease Control and Prevention.

Dr. Moon-Woo Seong has received research support from Korea Health Technology R&D Project, Korea Health Industry Development Institute (KHIDI), Ministry of Health & Welfare, Republic of Korea.

References

1. Cook Shukla L, Schulze J, Farlow J, Pankratz ND, Wojcieszek J, Foroud T. Parkinson Disease Overview. In: Adam MP, Ardinger HH, Pagon RA, et al., eds. GeneReviews((R)) Seattle (WA)1993.
2. Kalia LV, Lang AE. Parkinson's disease. Lancet 2015;386(9996):896–912. [PubMed: 25904081]
3. Kruger R, Kuhn W, Muller T, et al. Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease. Nat Genet 1998;18(2):106–108. [PubMed: 9462735]
4. Polymeropoulos MH, Lavedan C, Leroy E, et al. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science 1997;276(5321):2045–2047. [PubMed: 9197268]
5. Puschmann A, Jimenez-Ferrer I, Lundblad-Andersson E, et al. Low prevalence of known pathogenic mutations in dominant PD genes: A Swedish multicenter study. Parkinsonism Relat Disord 2019.
6. Tan MMX, Malek N, Lawton MA, et al. Genetic analysis of Mendelian mutations in a large UK population-based Parkinson's disease study. Brain 2019;142(9):2828–2844. [PubMed: 31324919]
7. La Cognata V, Morello G, D'Agata V, Cavallaro S. Copy number variability in Parkinson's disease: assembling the puzzle through a systems biology approach. Hum Genet 2017;136(1):13–37. [PubMed: 27896429]
8. Tagliafierro L, Chiba-Falek O. Up-regulation of SNCA gene expression: implications to synucleinopathies. Neurogenetics 2016;17(3):145–157. [PubMed: 26948950]
9. Fuchs J, Nilsson C, Kachergus J, et al. Phenotypic variation in a large Swedish pedigree due to SNCA duplication and triplication. Neurology 2007;68(12):916–922. [PubMed: 17251522]
10. Carvalho CM, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. Nat Rev Genet 2016;17(4):224–238. [PubMed: 26924765]
11. Lee JA, Carvalho CM, Lupski JR. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell 2007;131(7):1235–1247. [PubMed: 18160035]
12. Hastings PJ, Ira G, Lupski JR. A microhomology-mediated break-induced replication model for the origin of human copy number variation. PLoS Genet 2009;5(1):e1000327. [PubMed: 19180184]

13. Ross OA, Braithwaite AT, Skipper LM, et al. Genomic investigation of alpha-synuclein multiplication and parkinsonism. Ann Neurol 2008;63(6):743–750. [PubMed: 18571778]
14. Cardoso AR, Oliveira M, Amorim A, Azevedo L. Major influence of repetitive elements on disease-associated copy number variants (CNVs). Hum Genomics 2016;10(1):30. [PubMed: 27663310]
15. Mutez E, Lepretre F, Le Rhun E, et al. SNCA locus duplication carriers: from genetics to Parkinson disease phenotypes. Hum Mutat 2011;32(4):E2079–2090. [PubMed: 21412942]
16. Ahn TB, Kim SY, Kim JY, et al. alpha-Synuclein gene duplication is present in sporadic Parkinson disease. Neurology 2008;70(1):43–49. [PubMed: 17625105]
17. Shin CW, Kim HJ, Park SS, Kim SY, Kim JY, Jeon BS. Two Parkinson's disease patients with alpha-synuclein gene duplication and rapid cognitive decline. Mov Disord 2010;25(7):957–959. [PubMed: 20222138]
18. Raczy C, Petrovski R, Saunders CT, et al. Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. Bioinformatics 2013;29(16):2041–2043. [PubMed: 23736529]
19. Boeva V, Popova T, Bleakley K, et al. Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics 2012;28(3):423–425. [PubMed: 22155870]
20. Han W, Liu Y, Mi Y, Zhao J, Liu D, Tian Q. Alpha-synuclein (SNCA) polymorphisms and susceptibility to Parkinson's disease: a meta-analysis. Am J Med Genet B Neuropsychiatr Genet 2015;168B(2):123–134. [PubMed: 25656566]
21. Zhang Y, Shu L, Sun Q, Pan H, Guo J, Tang B. A Comprehensive Analysis of the Association Between SNCA Polymorphisms and the Risk of Parkinson's Disease. Front Mol Neurosci 2018;11:391. [PubMed: 30410434]
22. Ameur A, Dahlberg J, Olason P, et al. SweGen: a whole-genome data resource of genetic variability in a cross-section of the Swedish population. Eur J Hum Genet 2017;25(11):1253–1260. [PubMed: 28832569]
23. Bacolla A, Tainer JA, Vasquez KM, Cooper DN. Translocation and deletion breakpoints in cancer genomes are associated with potential non-B DNA-forming sequences. Nucleic Acids Res 2016;44(12):5673–5688. [PubMed: 27084947]
24. Beveridge DL, Dixit SB, Barreiro G, Thayer KM. Molecular dynamics simulations of DNA curvature and flexibility: helix phasing and premelting. Biopolymers 2004;73(3):380–403. [PubMed: 14755574]
25. Singh GB, Kramer JA, Krawetz SA. Mathematical model to predict regions of chromatin attachment to the nuclear matrix. Nucleic Acids Res 1997;25(7):1419–1425. [PubMed: 9060438]
26. Beck CR, Carvalho CMB, Akdemir ZC, et al. Megabase Length Hypermutation Accompanies Human Structural Variation at 17p11.2. Cell 2019;176(6):1310–1324 e1310. [PubMed: 30827684]
27. Vissers LE, Bhatt SS, Janssen IM, et al. Rare pathogenic microdeletions and tandem duplications are microhomology-mediated and stimulated by local genomic architecture. Hum Mol Genet 2009;18(19):3579–3593. [PubMed: 19578123]
28. Carvalho CM, Zhang F, Lupski JR. Structural variation of the human genome: mechanisms, assays, and role in male infertility. Syst Biol Reprod Med 2011;57(1–2):3–16. [PubMed: 21210740]
29. Bacolla A, Wells RD. Non-B DNA conformations, genomic rearrangements, and human disease. J Biol Chem 2004;279(46):47411–47414. [PubMed: 15326170]
30. Toffolatti L, Cardazzo B, Nobile C, et al. Investigating the mechanism of chromosomal deletion: characterization of 39 deletion breakpoints in introns 47 and 48 of the human dystrophin gene. Genomics 2002;80(5):523–530. [PubMed: 12408970]
31. Wells RD. Non-B DNA conformations, mutagenesis and disease. Trends Biochem Sci 2007;32(6):271–278. [PubMed: 17493823]
32. Marras C, Lang A, van de Warrenburg BP, et al. Nomenclature of genetic movement disorders: Recommendations of the international Parkinson and movement disorder society task force. Mov Disord 2016;31(4):436–457. [PubMed: 27079681]
33. Zafar F, Valappil RA, Kim S, et al. Genetic fine-mapping of the Iowan SNCA gene triplication in a patient with Parkinson's disease. NPJ Parkinsons Dis 2018;4:18. [PubMed: 29928688]

34. Coutelier M, Burglen L, Mundwiller E, et al. GRID2 mutations span from congenital to mild adult-onset cerebellar ataxia. Neurology 2015;84(17):1751–1759. [PubMed: 25841024]
35. Nussbaum RL. Genetics of Synucleinopathies. Cold Spring Harb Perspect Med 2018;8(6).
36. Nishioka K, Hayashi S, Farrer MJ, et al. Clinical heterogeneity of alpha-synuclein gene duplication in Parkinson's disease. Ann Neurol 2006;59(2):298–309. [PubMed: 16358335]

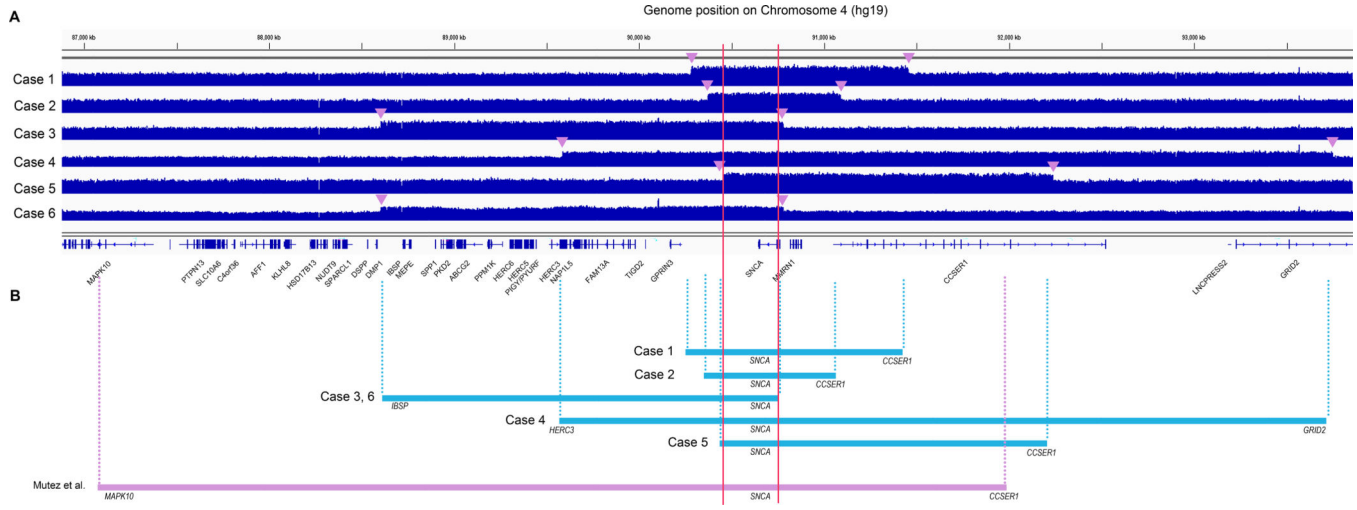


Figure 1.

Whole genome sequencing data for 6 patients with the confirmed duplication. (A) Duplicated regions shown by tiled data file format in IGV. Breakpoints are indicated by arrowheads. (B) Duplicated regions are denoted by segments showing the critical region only containing *SNCA* (red line).

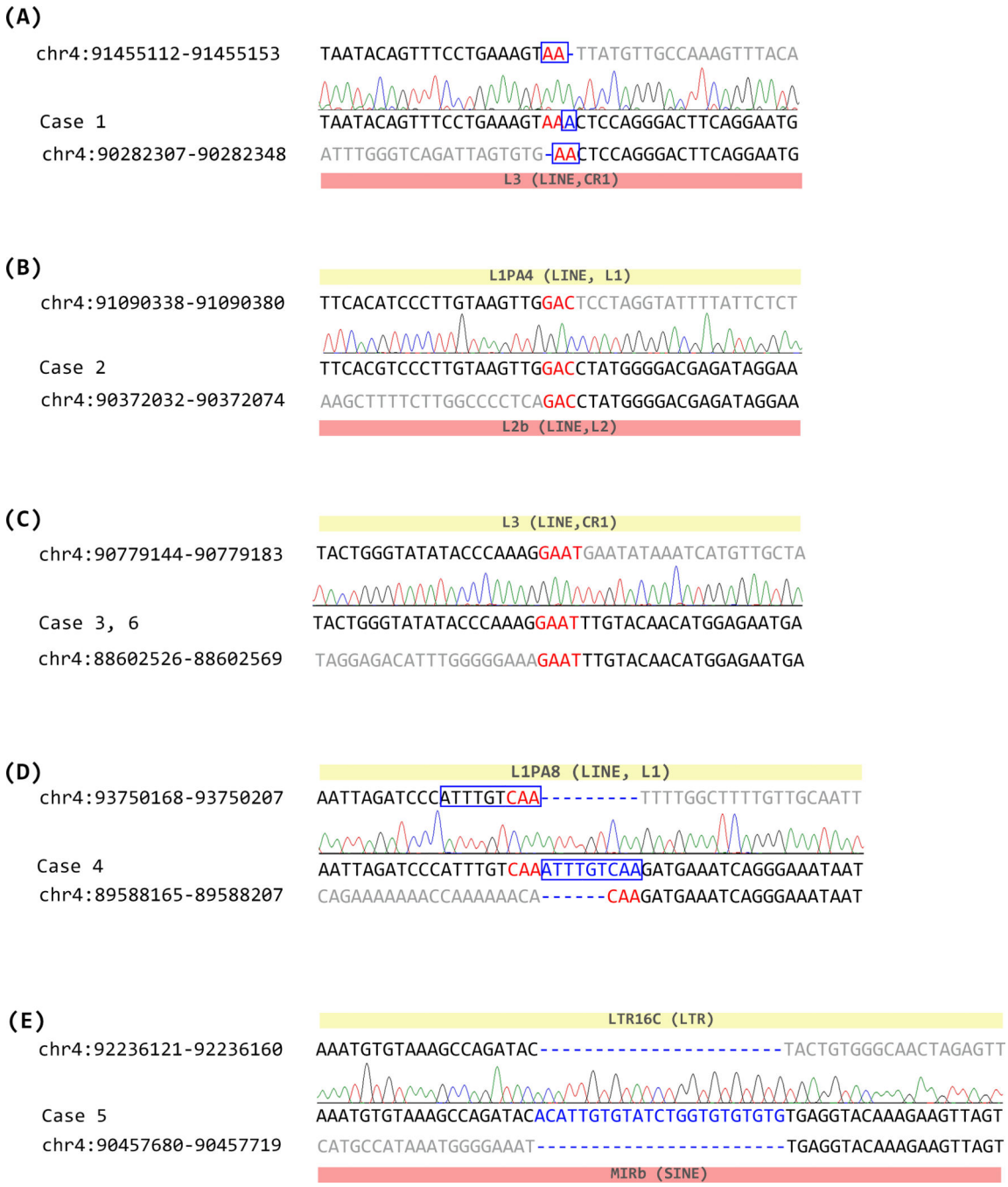
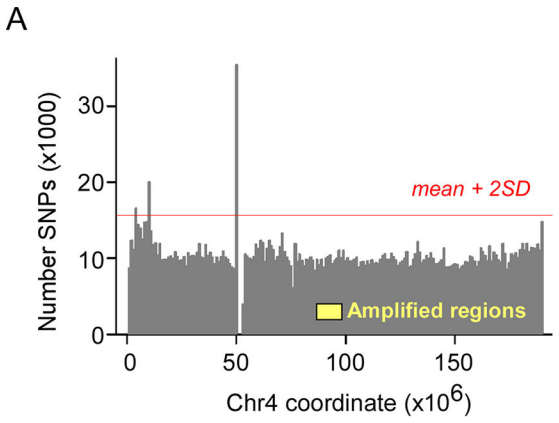


Figure 2.

Junction sequences of *SNCA* duplication cases. Nucleotide positions from chromosome 4 (GCRh37/hg19) are indicated. Original sequence of the distal breakpoint, junction sequence, and original sequence of the proximal breakpoint are represented for each case. Sequences downstream of the distal breakpoint or upstream of the proximal breakpoint are indicated in gray. Short stretch of microhomology is shown in red, and nucleotide insertions at the duplication junctions are shown in blue. Homologous region found near the breakpoints are indicated with a blue box. Repetitive elements found at a distal breakpoint are indicated by a

yellow box, and those on a proximal breakpoint are indicated with a red box. (A) junction sequence of Case 1, (B) junction sequence of Case 2, (C) junction sequence of Cases 3 and 6, (D) junction sequence of Case 4, (E) junction sequence of Case 5.



B

PONDS/100bp

	Cases	Controls	P-value
Direct repeats	0.0200	0.0248	NS
Inverted repeats	0.0600	0.1064	NS
G4 DNA	0.0100	0.0123	NS
Triplex DNA	0.0600	0.1227	NS
Z-DNA	0.0200	0.0112	NS

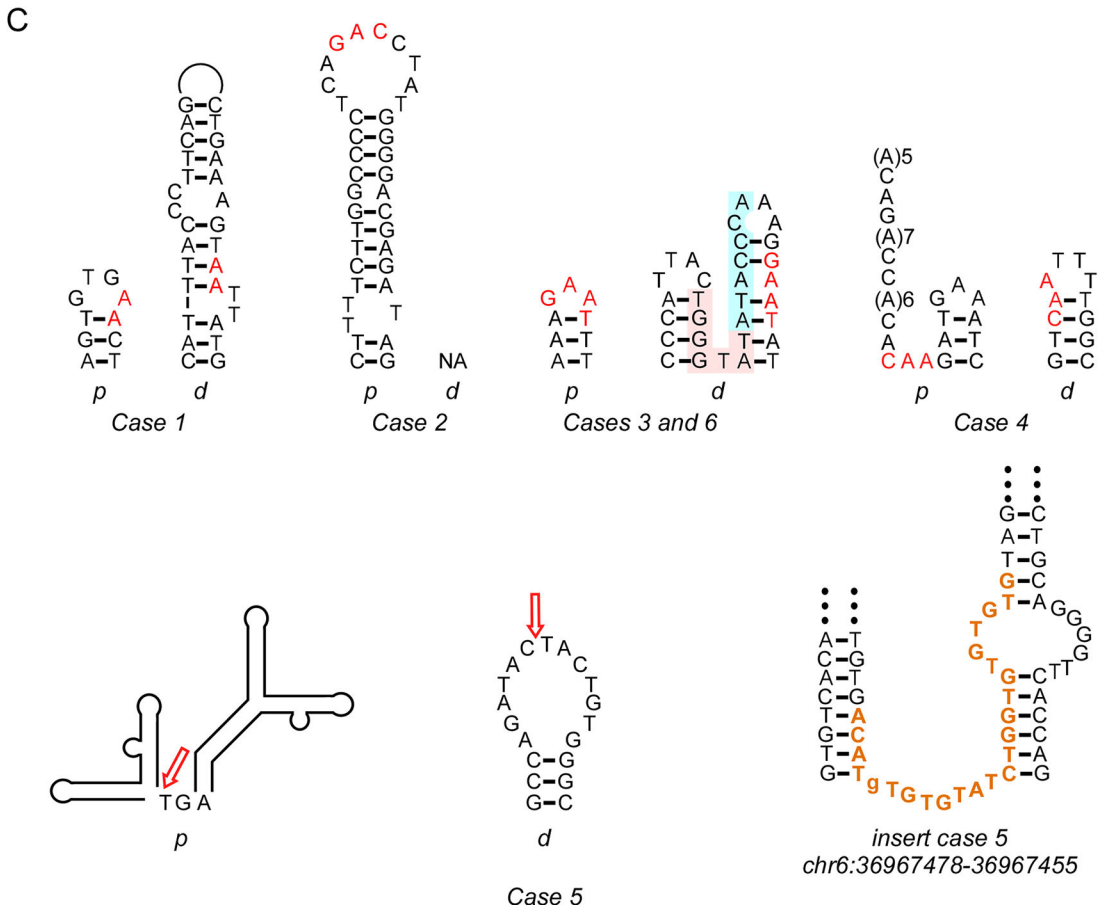


Figure 3.

Duplication junctions occur at potential stem-loop structures. (A) Histogram of the number of SNPs on chromosome 4; yellow, approximate interval of *SNCA* duplications; red line, upper limit for the number of SNPs (mean + 2 SD). (B) PONDS densities in cases and controls; p-values from z-tests. (C) Predicted stem-loop structures from Mfold at junctions; red, microhomologies; pink and blue highlight, additional inverted repeat; (A)n, sites of

predicted static DNA curvature; arrows, junctions; bold maroon, microinsertion matching a motif on the minus strand of chr6 and chr21; p, proximal, d, distal.

Table 1.

Genomic characteristics of the duplicated regions of chromosome 4

Case	Coordinates of duplicated region	Genes included in the duplicated region	Size (kb)
#1	chr4:90282327–91455133	*SNCA, MMRN1, CCSER1**	1,172.8
#2	chr4:90372052–91090360	*SNCA, MMRN1, CCSER1**	718.3
#3	chr4:88602546–90779167	*IBSP, MEPE, SPP1, PKD2, ABCG2, PPM1K, HERC6, HERC5, PIGY, PYURF, HERC3, NAP1L5, FAM13A, TIGD2, GPRIN3, SNCA*	2,176.6
#4	chr4:89588185–93750187	*HERC3**, *NAP1L5, FAM13A, TIGD2, GPRIN3, SNCA, MMRN1, CCSER1, LNCPRESS2, GRID2**	4,162.0
#5	chr4:90457700–92236140	*SNCA, MMRN1, CCSER1**	1,778.4
#6	chr4:88602546–90779167	*IBSP, MEPE, SPP1, PKD2, ABCG2, PPM1K, HERC6, HERC5, PIGY, PYURF, HERC3, NAP1L5, FAM13A, TIGD2, GPRIN3, SNCA*	2,176.6
Mutez, et al. (2011)	chr4:87172377–91943911	*MAPK10**, *PTPN13, SLC10A6, C4orf36, AFF1, KLHL8, HSD17B13, NUDT9, SPARCL1, DSPP, DMP1, IBSP, MEPE, SPP1, PKD2, ABCG2, PPM1K, HERC6, HERC5, PIGY, PYURF, HERC3, NAP1L5, FAM13A, TIGD2, GPRIN3, SNCA, MMRN1, CCSER1**	4,771.5

* Genes localized in the breakpoints

Table 2.

Clinical features of Korean patients with *SNCA* duplication

Case		#1	#2	#3	#4	#5	#6
Sex		M	M	F	F	F	M
Age of onset		40	51	66	55	49	51
Family history		Y	N*	N*	N	N	N
Initial symptoms	**Bradykinesia**	Y	Y	Y	NA	Y	Y
	Rigidity	Y	Y	Y	Y	NA	NA
	Tremor	Y	Y	Y	Y	Y	Y
	Postural instability	Y	Y	NA	NA	NA	NA
Disease Severity	**Disease duration (yr) at the time of evaluation**	1	2	2	5	4	4
	UPDRS motor	32	31	32	36	NA	NA
	H&Y scale	3	3	3	3	4	2
Nonmotor symptoms	**RBD**	NA	NA	NA	Y	Y	Y
	Cognition	Y†	N (MMSE 26 at age 65)	Y†	Y (MMSE 18 at age 60)	Y (MMSE 8 at age 56)	Y (MMSE 11 at age 56)
	Urinary.Symptoms	Y	NA	NA	NA	NA	Y
	Orthostatic hypotension	Y	Y	Y	Y	Y	N
	Depression	NA	NA	Y	Y	NA	Y
	Constipation	NA	Y	Y	Y	NA	Y
	Other NMS	NA	anxiety	insomnia	NA	NA	NA
Clinical features in disease course	**Levodopa response**	Y	Y	Y	NA	Y	NA
	Wearing-off	Y	Y	NA	NA	Y	Y
	Dyskinesia	Y	Y	NA	NA	Y	NA
	Hallucination	Y	NA	NA	Y	Y	Y
	Freezing of Gait	NA	Y	NA	NA	Y	NA
Pyramidal symptoms		Knee Jerk 3+/3+, Babinski sign (+/+)	NA	NA	NA	NA	NA

Case		#1	#2	#3	#4	#5	#6
Axial symptoms		Dysphonia	NA	NA	NA	NA	NA
EOM		cogwheel pursuit	DBN (Vib, HS)	NA	Hypometric saccade	NA	ocular flutter, hypometric saccade
Brain Imaging	**Brain MR**	Y(+, cortical atrophy in P-O)	NA	Y(−)	Y(−)	NA	Y(+−, mild cerebellar atrophy)
	CIT.PET	NA	Y(+)	NA	NA	NA	NA
	FDG.PET	NA	NA	NA	NA	Y(+,bilateral F-T-P decreased)	NA

*Asymptomatic carrier among family members (#2: Brother, #3: Daughter)

†MMSE scores were not available. A description of mild cognitive decline was found on the medical record.

Abbreviation: Y, symptom present; N, symptom not present; NA, not available; UPDRS, the Unified Parkinson's Disease Rating Scale; H&Y, Hoehn and Yahr scale; RBD, Rapid eye movment sleep behavior disorder; MMSE, Mini-Mental State Examination; NMS, nonmotor symptoms; DBN (HS, Vib); downbeat nystagmus by head shaking or vibration stimuli; P-O, parietal, occipital lobes; F-T-O, frontal, temporal, occipital lobes