Published in final edited form as:

Hum Mutat. 2013 August ; 34(8): 1071–1074. doi:10.1002/humu.22344.

High-Resolution Survey in Familial Parkinson Disease Genes Reveals Multiple Independent Copy Number Variation Events in *PARK2*

Liyong Wang^{1,2,*}, Karen Nuytemans¹, Guney Bademci¹, Cherylyn Jauregui¹, Eden R. Martin^{1,2}, William K. Scott^{1,2}, Jeffery M. Vance^{1,2}, and Stephan Zuchner^{1,2}

¹John P. Hussman Institute for Human Genomics, University of Miami, Miller School of Medicine, Miami, Florida

²Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miller School of Medicine, Miami, Florida

Abstract

A high density comparative genomic hybridization array was designed to evaluate CNVs in the genomic region of six familial PD genes in 181 PD cases and 67 controls. No CNV was found in *PARK7*, *ATP13A2*, *PINK1*, and *LRRK2*. Intronic-only CNVs were found in *SNCA* and *PARK2* but were not associated with PD risk. A whole-gene duplication of *SNCA* was found in one case. The allele frequency of *PARK2* exonic CNV is significantly higher in cases than in controls ($P = 0.02$), higher in early-onset (AAO ≤ 40) than in late-onset cases ($P = 0.001$), and higher in familial than in sporadic cases ($P = 0.005$). Except for single exon 2 duplications, all *PARK2* exonic CNVs have different breakpoints, even when the same exon(s) were involved. In conclusion, except for *SNCA* and *PARK2*, CNVs are not a major contributing mechanism for the familial PD genes examined. The majority of *PARK2* exonic CNVs are not recurrent.

Keywords

CNV; CGH; Parkinson; *PARK2*

DNA mutations in *PARK2* (MIM #602544), *SNCA* (MIM #163890), *PARK7* (MIM #602533), *PINK1* (MIM #608309), *ATP13A2* (MIM #610513), and *LRRK2* (MIM #609007) have been reported in Mendelian forms of Parkinson disease (PD). In recent years, it has been increasingly recognized that DNA CNVs represent a major source of variations in the human genome and have been associated with several complex phenotypes. Previously, several studies have examined the association between exonic CNVs in the familial PD genes using PCR-based methods, that is, quantitative PCR (qPCR) and multiplex ligation-

dependent probe amplification (MLPA) [Kay et al., 2010; Marder et al., 2010; Nuytemans et al., 2009; Pankratz et al., 2009; Sun et al., 2006]. Most of these studies have focused on *PARK2*, with only one study simultaneously examining five familial PD genes in a Belgian population [Nuytemans et al., 2009]. These studies, however, have examined CNV events of exons only. As a result, information on introns and promoter regions where regulatory elements reside could have been missed. In addition, without intronic markers and DNA from family members, it is difficult to infer continuous multiexon duplications/deletions as *cis*- or *trans*-events. Finally, it is laborious to evaluate the boundaries of each CNV event using PCR-based methods. To comprehensively evaluate CNVs in the PD familial genes, we performed a high-resolution comparative genomic hybridization (CGH) array screening of the genomic regions of six familial PD genes. Subsequently, we confirmed any potentially PD-associated CNVs using PCR-based methods.

The subjects used in this study were collected by the Morris K. Udall Parkinson Disease Research Center of Excellence (PDRCE) family ascertainment core at both Duke University and then at the University of Miami (J.M. Vance, PI), and the 13 centers of the Parkinson Disease Genetics Collaboration as described previously [Butler et al., 2011]. In total, 120 familial cases and 61 sporadic cases, and 67 unrelated controls were included in the study (Supp. Table S1). All subjects are non-Hispanic Caucasian. The genomic regions encompassing the six PD genes and their flanking areas (5 kb at the 5' end and 3 kb at the 3' end) were used as the target regions for probe design (Supp. Table S2). Custom 4 × 72 k format CGH microarrays were designed and manufactured by NimbleGen (Roche, Madison, WI, USA) with a probe spacing of 150 bp. A DNA pool of 16 members of CEPH families (Coriell, Camden, NJ, USA; 1:1 male/female ratio) was used as the reference sample. CGH array were processed following manufacturer's instructions. The CGH data were normalized with NimbleScan 2.4 and segmented with Nexus 4.0 (BioDiscovery, El Segundo, CA, USA) using the rank segmentation algorithm. Samples with excessive variation between probes were discarded (Supp. Methods). For CNV calling, the significance threshold for segmentation was set at 1×10^{-6} and the minimal number of probes required per segment was 50. Single copy number gains and losses were defined with a $\log_2$ Cy3/Cy5 ratio of 0.2 and -0.2, respectively. To confirm CNVs detected by the CGH arrays, we employed two PCR-based methods to determine the copy number of the target regions independently: MLPA (MRC-Holland, Amsterdam, Netherlands) or Taqman[®] CNV assays (AppliedBiosystems, Carlsbad, CA, USA). The reference sample used in the CGH experiments (pool of 16 CEPH individuals) was included in the MLPA and Taqman[®] CNV assays as the calibrator.

A one-tailed Fisher's exact *t*-test was used to compare CNV frequency in cases versus controls, early-onset cases (AAO ≤ 40) versus late-onset cases (AAO > 40) and familial cases versus sporadic cases, under the hypothesis that CNVs are more frequent in cases, especially early-onset and familial cases. All other frequency comparisons were evaluated using the two-tailed Fisher's exact *t*-test.

In total, 248 individual samples were screened with the customized CGH array. No CNVs were observed in *PARK7* (Supp. Table S3). Whole gene duplications of *PINK1*, whole gene duplications and deletions of *ATP13A2*, and exonic deletions in *LRRK2* were observed in

different individuals. In *SNCA* and *PARK2*, both exonic- and intronic-only CNVs were observed. The frequency of the intronic-only CNVs in these two genes, however, was not significantly different between cases and controls (Supp. Table S3). Therefore, we focused our follow-up effort on confirming exonic CNVs. None of the CNVs in *PINK1*, *ATP13A2*, and *LRRK2* could be confirmed with PCR-based methods, suggesting these observations are artifacts of the CGH arrays. In *SNCA*, the CGH analysis identified a whole gene duplication in one PD case with AAO of 36 y, which was confirmed with a Taqman[®] CNV assay targeting exon 4. The patient has a positive family history of PD with three affected persons in three generations. Only two members of the family were available for DNA testing: the proband and his unaffected sibling. The *SNCA* duplication was absent in the unaffected sibling. In *PARK2*, the CGH analysis identified CNVs involving one or multiple exons in 15 cases and two controls. All *PARK2* exonic CNV events were confirmed using MLPA and/or Taqman[®] CNV assays (Supp. Table S4) except in one control, in which neither MLPA nor Taqman[®] assays targeting the inside of the CNV demonstrated a copy number change. The absence of validation is unlikely because of technical faults as a positive control sample worked for both the MLPA and Taqman assay. The one control with confirmed *PARK2* exon 2 duplication has a family history of Alzheimer disease, although he/she has no PD family history or symptoms when he/she was examined at the age of 71 y. Table 1 displays the frequencies of *PARK2* exonic and intronic CNVs in different groups. Exonic CNVs are significantly more frequent in cases than in controls (5.6% vs. 0.8%, $P = 0.01$). Within the cases, they are significantly more frequent in early-onset cases (AAO ≤ 40) versus late-onset cases (12.0% vs. 2.6%, $P = 0.001$) as well as in familial cases versus sporadic cases (8.0% vs. 0.9%, $P = 0.005$). On the contrary, intronic CNVs are not significantly different between cases and controls, nor different case groups.

The *PARK2* gene has 12 exons, of which exons 2–9 are involved in CNV events in our study. Single exonic duplications and deletions of exon 2 were observed in seven cases and one control, accounting for almost half of the observed exonic CNV events in *PARK2* (Fig. 1 and Supp. Table S5). The smallest *PARK2* exonic CNV encompasses only exon 2 extending 9.4 kb in size, whereas the biggest *PARK2* exonic CNV spans exon 2–4 in a total deleted region of 450 kb (Fig. 1). All the exonic CNVs have different breakpoints except for the exon 2 duplications, which located in the same general region with similar breakpoints (Fig. 1). The exact breakpoints of the duplications, however, remain elusive. CNVs involving multiple exons (both continuous and discontinuous) were found in four cases and they are all heterozygous deletions (Supp. Fig. S1). CNVs involving multiple continuous exons also have the introns in between them deleted, suggesting that they are all *cis*-events. Thus, we did not perform additional segregation analysis to determine the *cis*- or *trans*-status of these CNVs. On the other hand, we performed segregation analysis on the two patients with discontinuous multiexon CNVs: Fam6_1 (deletion of exon 2 and exons 7–9) and Fam12_1 (deletion of exons 2–4 and exon 6) (Fig. 1) to determine whether the patients represent compound heterozygote carriers. Both patient Fam6_1 and patient Fam12_1 are early-onset cases (AAO = 25 and 26 y, respectively) with positive family history (Supp. Fig. S2). In the case of patient Fam6_1, DNA was available for testing for five siblings (four unaffected and one affected) and the parents (both unaffected). The parents are heterozygous carriers of the exon 2 deletion and the exons 7–9 deletion, respectively. Both affected

individuals in the family, patient Fam6_1 and her sib (AAO 44 y), are compound heterozygote carriers of the exon 2 and exons 7–9 deletion. None of the parents, the heterozygote or wild type carriers amongst the siblings had developed PD when they were examined at the age of 56 y or older. For patient Fam12_1, although there are three affected persons within three generations, DNA was available only for the parents and two siblings of the proband, all unaffected. The deletion of exons 2–4 and the deletion of exon 6 were transmitted to the proband from his father and mother, respectively, whereas both siblings carried wild type alleles.

PARK2 is located within a common fragile site FRA6E accompanied with high degree of genomic instability and rearrangement [Mitsui et al., 2010]. Consistent with its genomic location, the CGH data implicated numerous intronic-only CNV events in *PARK2* (~18% in the studied population). These intronic-only CNVs, however, do not appear to be associated with PD risk collectively. Because of the limited sample size in the current study, we cannot exclude the possibility that some *PARK2* intronic CNVs, if not all, contribute to the susceptibility of PD. Larger-scale studies are needed to characterize these intronic CNVs in depth. On the other hand, *PARK2* exonic CNVs are significantly more frequent in cases than in controls, especially in early-onset and familial cases. Our CGH data demonstrate that the majority of the *PARK2* exonic CNVs are independent events with different sizes. This is consistent with a recent report where only one recurrent event (deletion of exon 3) was observed in two unrelated individuals after screening 67 European patients with AR-JP (autosomal recessive juvenile parkinsonism) [Mitsui et al., 2010]. The recurrence rate, however, appears to be much higher in the Asian patients with AR-JP (about 50%) [Mitsui et al., 2010]. In our study, exon 2 is the most frequently duplicated/deleted exon in *PARK2*. This high rearrangement rate associated with exon 2 is likely to be attributable to the high density of Alu elements in intron 2 [Periquet et al., 2001].

As *PARK2* is known as a recessive gene for juvenile PD, it remains debatable, yet of clinical importance, whether *PARK2* heterozygous mutation carriers are at higher risk to develop PD. Some suggested that heterozygous exonic CNVs but not point sequence mutations in *PARK2* are associated with increased PD risk [Pankratz et al., 2009], whereas others suggested that neither heterozygous CNVs nor point sequence mutations increase the risk for PD development [Kay et al., 2010]. Although the current study was not designed to answer this question, segregation in families with discontinuous multiexon CNVs reported here supports that heterozygous *PARK2* exonic CNV is not a strong risk factor for PD. All individuals carrying two alleles of *PARK2* exonic CNVs ($N = 3$) developed PD before age of 45 y, whereas all the family members carrying one copy of *PARK2* exonic CNV ($N = 6$) did not develop PD. In our case–control cohort, the allele frequency of *PARK2* exonic CNV in cases is not significantly different from the controls after removing the homozygotes and compound heterozygotes. However, our study is limited by the moderate sample size, especially the small number of controls. To fully address this question, a large set of samples with both CNV and sequence mutation data in *PARK2* is needed.

Several studies have estimated the frequency of *PARK2* exonic CNVs in PD patients, including PROGENI, GenePD, CORE-PD, NGRC, and a Belgian PD study [Kay et al., 2010; Marder et al., 2010; Nuytemans et al., 2009; Pankratz et al., 2009; Sun et al., 2006].

The frequency of *PARK2* exonic CNVs in our study is significantly higher than in others ($P < 0.0005$), except for the GenePD study (Supp. Table S6). This can most likely be ascribed to the enrichment for PD cases with positive family history and earlier AAO in our study. The PD patient groups in the CORE-PD study, the Belgian PD study, and the NGRC study are primarily composed of sporadic PD cases. The frequency of *PARK2* exonic CNVs in PD patients is around 1.3% in all three studies. This is lower than the frequencies in the PROGENI study (2.6%), the GenePD study (4.1%), and our study (5.6%), whose patient groups are primarily composed of familial PD cases, suggesting that the *PARK2* exonic CNVs are more common in familial than in sporadic cases. Interestingly, the frequency of *PARK2* exonic CNVs in the PROGENI study is significantly lower than in our study ($P < 0.0001$) despite PROGENI examining only familial cases. The cases in the PROGENI study, however, have relatively late AAO: median AAO of 60 y in PROGENI compared with mean AAO of 48 y in our study. This suggests that *PARK2* exonic CNVs are more common in PD cases with younger AAO. This notion is further supported by the GenePD study where familial PD cases with AAO <54 were examined; the frequency of *PARK2* exonic CNV in this cohort with younger AAO is 4.1%. It is higher than observed in the PROGENI study and close to our observation.

In addition to *PARK2* exonic CNVs, a whole gene duplication of *SNCA* was identified by CGH and confirmed by qPCR in a PD patient. Triplications and duplications of the entire *SNCA* gene have been reported in both familial and sporadic cases [Ahn et al., 2008; Nuytemans et al., 2010; Singleton et al., 2003]. The PD patient in our study has a positive family history with three affected persons in three generations; the proband and his maternal aunt and grandmother. Besides the proband, only his unaffected sibling has DNA available for testing and does not carry *SNCA* duplication. It is likely that the proband inherited the duplicated *SNCA* allele from his grandmother through his mother, suggesting the latter as an asymptomatic *SNCA* duplication carrier. This is consistent with the previous report identifying *SNCA* as an autosomal dominant gene for PD with incomplete penetrance of the gene duplication [Nishioka et al., 2006],

CNV genotyping is technically more challenging than SNP genotyping, partially because of the quantitative rather than qualitative nature of the CNV assay. Different data processing strategies and calling algorithms have been developed for CNV studies. No single method, however, can be considered the golden standard and validation is essential. In this study, we used CGH array in the screening stage and further examined any potential disease-associated CNVs using PCR-based methods. To minimize the chance for false negative findings, we employed a rather permissive setting to make CNV calls at the screening stage (Supp. Methods). On the basis of 171 samples screened by both CGH and SNP-array, the false negative rate is 0 with our CNV calling algorithm. On the other hand, our CGH analysis nominated a few false positive findings. Employing a more stringent calling algorithm will reduce these false positive findings but at a price of inflated false negative rates. For example, when the threshold of $\log_2$ Cy3/Cy5 ratio was increased from 0.2 to 0.3 in defining copy number gains, the whole-gene duplication of *SNCA*, which has been confirmed with q-PCR, would be missed. This highlights the challenge of CNV calling and emphasizes the importance of validation in the study of CNVs. Comparing with other

methods to detect CNVs, CGH has the highest initiating cost (~\$180 per sample for 72 thousands of probes, which are enough to cover several genes) and requires the most DNA (1 μ g) for one reaction. MLPA has the lowest initiating cost (~\$1.8 per exon) and requires 50–200 ng of DNA per reaction. Taqman qPCR requires the least amount of DNA (10–50 ng per reaction) and costs ~\$6.4 per exon (running assay in triplicate). Lately, a new technology NanoString has been developed to measure CNVs, which costs about \$5.0 per exon (assuming two probes per exon) and requires 300–500 ng of DNA per reaction, which can accommodate up to 800 probes. Among these methods, Taqman qPCR has the shortest assay time (~2.5 hr) comparing with others (2–3 days). The CGH-array technology provides the most comprehensive survey of interested region attributable to the high-density probes in each reaction, and therefore provides estimation on the size and breakpoints of CNVs. One disadvantage of CGH is that it is not sensitive to small structural rearrangement. The smallest CNV that can be detected with our CGH calling algorithm (50 probes per CNV) and probe density (150 bp spacing) is 7.5 kb. Because of potential noise introduced by ultra-high density probes, it is not recommended to go below 150 bp spacing in a CGH design. Using less probes to define a CNV event could increase the CNV calling resolution but the relaxed calling algorithm could lead to higher false positive rate.

In summary, our study suggests that except for the whole gene copy number gain of *SNCA* and *PARK2* exonic duplications/deletions, CNVs are not a major contributing mechanism for disease development in the other familial PD genes. This is consistent with a recent report on CNVs using the SNP-array in a large cohort of PD patients and controls, where no CNVs were observed in the *PINK1*, *PARK7*, *LRRK2*, and *ATP13A2* genes [Pankratz et al., 2011]. Furthermore, we showed that most *PARK2* exonic CNVs in Caucasians are independent events with different sizes and breakpoints, even when the same exon(s) were involved. Our study provided the information needed for subsequent sequencing analysis to characterize the breakpoints of the CNVs. Finally, our CGH data confirmed several *PARK2* CNVs involving continuous exons as *cis*-structural variations encompassing exons and in-between introns. This information facilitates the determination of the heterozygous or compound heterozygous status of carriers, which is important to interpret the correlation between *PARK2* genotype and PD risk.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to the patients and control subjects who participated in this study. We thank the members of the PD Genetics Collaboration: Martha A. Nance, Ray L. Watts, Jean P. Hubble, William C. Koller, Kelly Lyons, Rajesh Pahwa, Matthew B. Stern, Amy Colcher, Bradley C. Hiner, Joseph Jankovic, William G Ondo, Fred H. Allen, Jr., Christopher G. Goetz, Gary W. Small, Donna Masterman, Frank Mastaglia, and Jonathan L. Haines who contributed subjects to this study. Some of the samples used in this study were collected while the Udall PDRCE was at Duke University.

Contract grant sponsors: NIH NINDS (2P50NS071674-02); James and Esther King/Biomedical Research Program (1KN 14).

References

- Ahn TB, Kim SY, Kim JY, Park SS, Lee DS, Min HJ, Kim YK, Kim SE, Kim JM, Kim HJ, Cho J, Jeon BS. Alpha-synuclein gene duplication is present in sporadic Parkinson disease. *Neurology*. 2008; 70:43–49. [PubMed: 17625105]
- Butler MW, Burt A, Edwards TL, Zuchner S, Scott WK, Martin ER, Vance JM, Wang L. Vitamin D receptor gene as a candidate gene for Parkinson disease. *Ann Hum Genet*. 2011; 75:201–210. [PubMed: 21309754]
- Kay DM, Stevens CF, Hamza TH, Montimurro JS, Zabetian CP, Factor SA, Samii A, Griffith A, Roberts JW, Molho ES, Higgins DS, Ganther S, Moses L, et al. A comprehensive analysis of deletions, multiplications, and copy number variations in PARK2. *Neurology*. 2010; 75:1189–1194. [PubMed: 20876472]
- Marder KS, Tang MX, Mejia-Santana H, Rosado L, Louis ED, Comella CL, Colcher A, Siderowf AD, Jennings D, Nance MA, Bressman S, Scott WK, et al. Predictors of parkin mutations in early-onset Parkinson disease: the consortium on risk for early-onset Parkinson disease study. *Arch Neurol*. 2010; 67:731–738. [PubMed: 20558392]
- Mitsui J, Takahashi Y, Goto J, Tomiyama H, Ishikawa S, Yoshino H, Minami N, Smith DI, Lesage S, Aburatani H, Nishino I, Brice A, 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; 87:75–89. [PubMed: 20598272]
- Nishioka K, Hayashi S, Farrer MJ, Singleton AB, Yoshino H, Imai H, Kitami T, Sato K, Kuroda R, Tomiyama H, Mizoguchi K, Murata M, et al. Clinical heterogeneity of alpha-synuclein gene duplication in parkinson's disease. *Ann Neurol*. 2006; 59:298–309. [PubMed: 16358335]
- Nuytemans K, Meeus B, Crosiers D, Brouwers N, Goossens D, Engelborghs S, Pals P, Pickut B, Van den Broeck M, Corsmit E, Cras P, De Deyn PP, et al. Relative contribution of simple mutations vs. copy number variations in five Parkinson disease genes in the Belgian population. *Hum Mutat*. 2009; 30:1054–1061. [PubMed: 19405094]
- Nuytemans K, Theuns J, Cruts M, Van Broeckhoven C. Genetic etiology of Parkinson disease associated with mutations in the SNCA PARK2 PINK1 PARK7 and LRRK2 genes: a mutation update. *Hum Mutat*. 2010; 31:763–780. [PubMed: 20506312]
- Pankratz N, Dumitriu A, Hetrick KN, Sun M, Latourelle JC, Wilk JB, Halter C, Doheny KF, Gusella JF, Nichols WC, Myers RH, Foroud T, et al. Copy number variation in familial Parkinson disease. *PLoS ONE*. 2011; 6:e20988. [PubMed: 21829596]
- Pankratz N, Wilk JB, Latourelle JC, DeStefano AL, Halter C, Pugh EW, Doheny KF, Gusella JF, Nichols WC, Foroud T, Myers RH, et al. Genomewide association study for susceptibility genes contributing to familial Parkinson disease. *Hum Genet*. 2009; 124:593–605. [PubMed: 18985386]
- Periquet M, Lucking C, Vaughan J, Bonifati V, Durr A, De Michele G, Horstink M, Farrer M, Illarioshkin SN, Pollak P, Borg M, Brefel-courbon C, et al. Origin of the mutations in the parkin gene in europe: exon rearrangements are independent recurrent events, whereas point mutations may result from founder effects. *Am J Hum Genet*. 2001; 68:617–626. [PubMed: 11179010]
- Singleton AB, Farrer M, Johnson J, Singleton A, Hague S, Kachergus J, Hulihan M, Peuralinna T, Dutra A, Nussbaum R, Lincoln S, Crawley A, et al. Alpha-synuclein locus triplication causes Parkinson's disease. *Science*. 2003; 302:841. [PubMed: 14593171]
- Sun M, Latourelle JC, Wooten GF, Lew MF, Klein C, Shill HA, Golbe LI, Mark MH, Racette BA, Perlmutter JS, Parsian A, Guttman M, et al. Influence of heterozygosity for parkin mutation on onset age in familial Parkinson disease: the GenePD study. *Arch Neurol*. 2006; 63:826–832. [PubMed: 16769863]

Figure 1.

Boundaries of *PARK2* exonic CNVs in different individuals. Each horizontal colored bar represents the boundaries of one *PARK2* exonic CNV event: copy number gain in green and copy number loss in red. The gene structure is displayed on the top of the figure with the vertical gray bar representing exons and each exon is numbered underneath the gray bar.

Table 1

Allele Frequency of *PARK2* CNVs in Different Groups

Category	Number of subjects	Exonic			Intronic		
		Number of events	Frequency	P value*	Number of events	Frequency	P value
Case-control dataset							
Case	169	19	5.6%		62	18.3%	
Control	66	1	0.8%	0.01	23	17.4%	0.46
Case only							
Age at onset							
40	54	13	12.0%		19	17.6%	
>40	115	6	2.6%	0.001	43	18.7%	0.86
Family history							
Positive	113	18	8.0%		37	16.4%	
Negative	56	1	0.9%	0.005	25	22.3%	0.18

P values are from one-tailed *t*-test for exonic CNVs, based on the hypothesis that case carries more exonic CNVs than control, and are from two-tailed *t*-test for the intronic CNVs to explore the effect of intronic CNVs on PD risk.