Integrated sequencing and array comparative genomic hybridization in familial Parkinson disease

Laurie A. Robak, MD, PhD,* Renqian Du, PhD, DDS,* Bo Yuan, PhD, Shen Gu, PhD, Isabel Alfradique-Dunham, MD, Vismaya Kondapalli, BS, Evelyn Hinojosoa, BS, Amanda Stillwell, BS, Emily Young, BA, Chaofan Zhang, BS, Xiaofei Song, PhD, Haowei Du, MS, Tomasz Gambin, PhD, Shalini N. Jhangiani, PhD, Zeynep Coban Akdemir, PhD, Donna M. Muzny, MS, Anusha Tejomurtula, MS, Owen A. Ross, PhD, Chad Shaw, PhD, Joseph Jankovic, MD, Weimin Bi, PhD, Jennifer E. Posey, MD, PhD, James R. Lupski, MD, PhD,† and Joshua M. Shulman, MD, PhD†

Neurol Genet 2020;6:e498. doi:10.1212/NXG.0000000000000498

Correspondence
Dr. Shulman
Joshua.Shulman@bcm.edu
or Dr. Lupski
jlupski@bcm.edu

Abstract

Objective
To determine how single nucleotide variants (SNVs) and copy number variants (CNVs) contribute to molecular diagnosis in familial Parkinson disease (PD), we integrated exome sequencing (ES) and genome-wide array-based comparative genomic hybridization (aCGH) and further probed CNV structure to reveal mutational mechanisms.

Methods
We performed ES on 110 subjects with PD and a positive family history; 99 subjects were also evaluated using genome-wide aCGH. We interrogated ES and aCGH data for pathogenic SNVs and CNVs at Mendelian PD gene loci. We confirmed SNVs via Sanger sequencing and further characterized CNVs with custom-designed high-density aCGH, droplet digital PCR, and breakpoint sequencing.

Results
Using ES, we discovered individuals with known pathogenic SNVs in GBA (p.Glu365Lys, p.Thr408Met, p.Asn409Ser, and p.Leu483Pro) and LRRK2 (p.Arg1441Gly and p.Gly2019Ser). Two subjects were each double heterozygotes for variants in GBA and LRRK2. Based on aCGH, we additionally discovered cases with an SNCA duplication and heterozygous intragenic GBA deletion. Five additional subjects harbored both SNVs (p.Asn52Metfs*29, p.Thr240Met, p.Pro437Leu, and p.Trp453*) and likely disrupting CNVs at the PRKN locus, consistent with compound heterozygosity. In nearly all cases, breakpoint sequencing revealed microhomology, a mutational signature consistent with CNV formation due to DNA replication errors.

Conclusions
Integrated ES and aCGH yielded a genetic diagnosis in 19.3% of our familial PD cohort. Our analyses highlight potential mechanisms for SNCA and PRKN CNV formation, uncover multilocus pathogenic variation, and identify novel SNVs and CNVs for further investigation as potential PD risk alleles.

*These authors contributed equally to this manuscript as co–first authors.
†These authors contributed equally to this manuscript as co–last authors.
Up to 20% of patients with Parkinson disease (PD) report a positive family history, and genetic risk factors are more common in these families. Identification of specific genetic risk factors can reveal prognostic information, such as risk of cognitive impairment and/or rate of progression, and may soon highlight eligibility for personalized therapies. In addition, discovery of risk variants may inform genetic counseling of unaffected family members. Indeed, surveys of patients with PD and caregivers reveal a high level of interest in genetic testing for PD.

More than 40 different loci that increase PD susceptibility have been identified in familial and sporadic PD. Exome sequencing (ES) is ideally suited to identify single nucleotide variants (SNVs) in genetically heterogeneous diseases. In a study of adult patients referred for diverse clinical indications, ES had a diagnostic yield of 10% in individuals older than 30 years. In a recent study of 80 early-onset sporadic PD cases, ES yielded an overall diagnostic rate of 11%, with GBA alleles accounting for 5%. Nevertheless, clinical genetic testing is not routinely performed for PD, and ES remains poorly studied as a potential genetic diagnostic tool.

Although most identified PD risk alleles are SNVs, chromosomal structural rearrangements, or copy number variants (CNVs), also play an important role. Despite notable recent advances, ES remains insensitive for detection of small CNVs (<50 kb). Several complementary approaches, including multiplex ligation-dependent probe amplification, bacterial artificial chromosome arrays, and single nucleotide polymorphism arrays, have shown mixed success for identification of CNVs in PD cohorts. In contrast, genome-wide array-based comparative genomic hybridization (aCGH) is a highly-validated, sensitive clinical screening tool for CNV detection, offering exon-by-exon coverage for a multitude of disease-associated genes. Although not yet adopted in most diagnostic laboratories, droplet digital PCR (ddPCR) is also emerging as a rapid and cost-efficient, targeted approach for the assessment of small CNVs at specific loci. Compared with standard quantitative PCR, digital PCR offers enhanced copy number and gene dosage sensitivity, precision, and reliability due to sample partitioning. In addition, mechanisms of CNV formation in PD remain understudied.

To our knowledge, integrated ES and aCGH for analysis of SNVs and CNVs, respectively, have not previously been systematically used in PD. We hypothesized that ES and aCGH in combination will yield an increased genetic molecular diagnostic rate. We also evaluated ddPCR as a novel strategy for confirmation of pathogenic CNVs in PD, and using breakpoint sequencing, we investigated potential mechanisms for CNV formation.

Methods

See supplementary information (links.lww.com/NXG/A305) for complete methods, including further details and references.

Subjects

We studied 110 PD cases evaluated in the Baylor College of Medicine (BCM) PD Center and Movement Disorders Clinic in Houston, TX, with a family history of PD. As a positive control for aCGH, we included a sample from a known subject with an SNCA triplication. We also interrogated a Baylor Genetics diagnostic laboratory sample including 12,922 clinical referral samples for aCGH from peripheral blood using either v9 or v10 Baylor arrays. Subject numbers throughout the text are consistent with clinical and demographic details provided in table e-1 (links.lww.com/NXG/A306).

Standard protocol approvals, registrations, and subject consents

All subjects provided informed consent. The BCM Institutional Review Board approved this study along with the analysis of aggregate clinical genomic data.

Gene set definition and variant criteria

We focused our analyses on genes and variants established to cause familial PD, including the autosomal dominant loci, SNCA (PARK1, MIM#168601), GBA (MIM#168600), LRRK2 (MIM#607060), GCH1 (MIM#600225), DNAJC13 (MIM#616361), and VPS35 (MIM#614203), as well as the autosomal recessive loci, PRKN (PARK2, MIM#600116), PINK1 (MIM#605909) and PARK7 (DJ1, MIM#606324), based on the available literature in April 2015 when this study was initiated. In our CNV analyses, we also considered deletions at 22q11.2. Gene names in this study conform to current guidelines from the HUGO Gene Nomenclature Committee (genenames.org). All pathogenic alleles included in this study are well-established, nonsynonymous coding variants with moderate to high penetrance (odds ratio [OR] >2) meeting stringent evidence for replication across studies or within the same study. We considered all other variants discovered in these genes but not previously reported in PD to be variants of unknown significance (VUSs).
Detection and confirmation of SNVs and CNVs

We extracted genomic DNA from peripheral blood samples obtained from each participant and performed ES using the Illumina HiSeq 2000 at the BCM Human Genome Sequencing Center. Samples achieved an average of 95% of targeted exome bases covered to a depth of 20X or greater. All pathogenic SNVs detected were confirmed via Sanger sequencing. Genome-wide array CGH was performed on 99 of 110 subjects for which sufficient DNA remained using Baylor Genetics v10 2x400K clinical-grade oligonucleotide microarrays. We defined potential CNVs as those regions with 3 or more consecutive probes with consistent direction of effect. For confirmation, we used a custom 8x60K high-density array through Agilent (Santa Clara, CA). To confirm CNVs at PRKN and GBA, we additionally performed ddPCR (see e-Methods for detailed protocol, links.lww.com/NXG/A305).

Data availability

For all subjects in the BCM cohort who consented to allow for public data sharing, ES and aCGH are in process for release in relevant genomic databases. The complete ES and aCGH data sets are also available on request by contacting the corresponding author, Dr. Shulman (joshua.shulman@bcm.edu).

Results

We pursued genetic diagnostic evaluation of 110 total subjects (including 109 unrelated probands) with familial PD. The mean age at onset was 50 years (SD = 15); 51% were male. The ethnic composition of the cohort was 72% Caucasian, 17% Hispanic, 6% East Asian, South Asian or Middle Eastern, and 6% undefined (not reported).

Single nucleotide variants

We first examined subject ES data for pathogenic SNVs in established PD genes (see Methods). Among the dominant PD loci, 15 individuals had variants in LRRK2 (c.6055G>A:p.Gly2019Ser and c.4321C>G:p.Arg1441Gly) and GBA (c.1093G>A:p.Glu365Lys, c.1223C>T:p.Thr408Met, c.1448T>C:p.Leu483Pro, and c.1226A>G:p.Asn409Ser) (table 1 and table e-1, links.lww.com/NXG/A306). Two subjects each harbored heterozygous SNVs in both GBA and LRRK2, i.e., were double heterozygotes. One such subject had a combination of LRRK2 p.Gly2019Ser and GBA p.Glu365Lys (subject 2), whereas the other had LRRK2 p.Gly2019Ser and GBA p.Leu483Pro (subject 13). Both subjects had onset of PD symptoms in their 40s. On initial examination, subject 2 had tremor at rest, rigidity, bradykinesia, and dystonic posturing in both hands. She reported a history of PD in her father and paternal grandfather (figure e-1A, links.lww.com/NXG/A305, and table e-1, links.lww.com/NXG/A306). There was no history of cognitive impairment or dementia. Subject 13 presented with resting tremor, rigidity, and bradykinesia. She reported a family history of PD in her paternal uncle (figure e-1B, links.lww.com/NXG/A305, and table e-1, links.lww.com/NXG/A306). Ten years after PD diagnosis, she developed visual hallucinations and delusions. The subjects were of European and Hispanic ancestry, respectively; neither reported Ashkenazi Jewish heritage.

ES also revealed 7 individuals with pathogenic variants in loci usually associated with autosomal recessive PD, including PRKN (6 individuals) and PARK7 (1 individual). However, all subjects were heterozygous SNV carriers, and therefore, isolated ES was nondiagnostic (table 1). Therefore, based on ES alone, we identified a pathogenic variant accounting for PD in 13.8% (n = 15 of 109 probands) of our familial PD cohort. We confirmed all implicated variants via Sanger sequencing. Besides the pathogenic variants noted above, ES also identified heterozygous VUSs in many PD risk genes (table e-2, links.lww.com/NXG/A306).

Copy number variants

We next interrogated aCGH data for pathogenic CNVs among PD genes. Our analyses included 99 of 110 total subjects evaluated by ES. We did not detect any CNVs in VPS35, LRRK2, DNAJC13, GCH1, PARK7, or PINK1, nor did we identify any candidate deletions at the 22q11.2 locus. However, we discovered CNVs in SNCA (n = 1), GBA (n = 1), and PRKN (n = 5). We confirmed all reported CNVs through custom high-density arrays and breakpoint sequencing. CNVs in SNCA and GBA affect dominant PD genes and were therefore diagnostic based on aCGH alone. We consider all heterozygous CNVs in the recessive PD gene, PRKN, in combination with ES results (see next section). Overall, isolated aCGH identified a diagnostic genetic risk factor for PD in 2.0% of our cohort (n = 2 of 99 probands). Based on aCGH, we also detected numerous large CNVs (>1 Mb) within our cohort that affect other genomic loci; these variants remain of uncertain clinical significance (table e-3, links.lww.com/NXG/A306).

In subject 3, we detected a 248-kb duplication encompassing SNCA, as well as the adjacent gene, MMRN1. We confirmed this CNV by high-density aCGH and breakpoint analysis (figure 1A). On initial examination, this subject exhibited rigidity, tremor, and gait impairment, along with hyperreflexia and clonus. The subject was of Hispanic and Native American ancestry; the subject’s father had PD with dementia. Besides providing independent confirmation, breakpoint sequencing can provide clues to mechanisms of CNV formation. In the case of subject 3, we identified a 1-bp microhomology domain, which is a short sequence that is identical to another region in the genome reduced from 2 copies to 1 during the template switch accompanying replicative repair. Microhomology is characteristic of certain DNA replication errors that can generate CNVs (see Discussion). As a positive control for our aCGH analysis, we also included a known SNCA triplication sample from the index family in which SNCA locus multiplication was first discovered as a cause for PD. Breakpoint sequencing revealed that this copy number alteration is a 1.7-Mb complex genomic rearrangement (figure 1B), consisting of a duplication-inverted triplication-duplication (DUP-TRP/INV-DUP). This finding confirms and extends prior investigation of this particular structural variant and is also consistent with a likely replication-based mechanism for CNV formation.
We also discovered a heterozygous intragenic deletion in GBA in 1 subject (figure 2), who presented at age 28 years with tremor, bradykinesia, and rigidity. She had an excellent response to levodopa in her early 30s and subsequently developed dyskinesia. The subject was of European descent, and both of her maternal grandparents were also diagnosed with PD (figure e-1C, links.lww.com/NXG/A305). Because variant confirmation at the GBA locus can be complicated by an adjacent pseudogene with significant homology, we confirmed the 4.7-kb deletion of exons 2–8 using long-range PCR. Using breakpoint sequencing, we also confirmed heterozygosity and further revealed a 5-bp microhomology domain consistent with CNV formation due to nonhomologous recombination or replication errors (figure 2B).

**Integrated analysis of SNVs and CNVs**

As highlighted above, isolated ES and aCGH each identified a number of subjects with heterozygous SNVs or CNVs affecting recessive genes. To determine whether these changes might be diagnostic, we next examined the results together for potential biallelic variation due to both an SNV allele and a CNV allele. Indeed, 3 subjects in our cohort were newly identified as potential compound heterozygous carriers of both a pathogenic CNV and SNV in PRKN (figure 3). Subject 6 had a 364-kb duplication of exons 4–6 and a frameshift deletion c.155delA:p.Asn52Metfs*29. Subject 20 had a 222-kb deletion of exons 8 and 9 and a stopgain c.1358G>A:p.Trp453*. Subject 11 harbored a pathogenic PRKN SNV (c.1310C>T:p.Pro437Leu) and a complex locus rearrangement, including a copy number neutral region flanked by 404 kb and 199 kb duplications affecting exons 2 and 3 and exons 5 and 6, respectively (figure 3B). Of interest, ES also discovered an additional VUS (c.2T>C:p.Met1Thr). Based on available clinical information (table e-1, links.lww.com/NXG/A306), all subjects with PRKN variants had young-onset PD (age range 15–36 years).

We again confirmed all CNVs using custom, high-density arrays as well as breakpoint sequencing. In the case of subject 6, we were unable to successfully amplify breakpoint junctions despite multiple attempts, suggesting a more complex genomic rearrangement or raising the possibility that this duplication is located elsewhere in the genome. For all other PRKN CNVs, figure 3 shows the junction structures, highlighting likely mechanisms of CNV formation. Overall, integrated ES and CNV identified a genetic cause for PD in 4 additional

<table>
<thead>
<tr>
<th>Gene</th>
<th>Variant Description</th>
<th>Diagnostic</th>
<th>Subjects (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dominant</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LRRK2</td>
<td>c.6055G&gt;A:p.Gly2019Ser</td>
<td>Y</td>
<td>3&lt;sup&gt;a,b&lt;/sup&gt;</td>
</tr>
<tr>
<td>LRRK2</td>
<td>c.4321C&gt;G:p.Arg1441Gly</td>
<td>Y</td>
<td>1</td>
</tr>
<tr>
<td>GBA</td>
<td>c.1093G&gt;A:p.Glu365Lys</td>
<td>Y</td>
<td>5&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>GBA</td>
<td>c.1223C&gt;T:p.Thr408Met</td>
<td>Y</td>
<td>5</td>
</tr>
<tr>
<td>GBA</td>
<td>c.1448T&gt;C:p.Leu483Pro</td>
<td>Y</td>
<td>1&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>GBA</td>
<td>c.1226A&gt;G:p.Asn409Ser</td>
<td>Y</td>
<td>2</td>
</tr>
<tr>
<td>Recessive</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRKN</td>
<td>c.155delA:p.Asn52Metfs*29</td>
<td>Y&lt;sup&gt;*&lt;/sup&gt;</td>
<td>1</td>
</tr>
<tr>
<td>PRKN</td>
<td>c.1310C&gt;T:p.Pro437Leu&lt;sup&gt;4&lt;/sup&gt;</td>
<td>Y&lt;sup&gt;*&lt;/sup&gt;</td>
<td>1</td>
</tr>
<tr>
<td>PRKN</td>
<td>c.1358G&gt;A:p.Trp453*</td>
<td>Y&lt;sup&gt;*&lt;/sup&gt;</td>
<td>1</td>
</tr>
<tr>
<td>PRKN</td>
<td>c.719C&gt;T:p.Thr240Met</td>
<td>Y&lt;sup&gt;*&lt;/sup&gt;</td>
<td>2</td>
</tr>
<tr>
<td>PRKN</td>
<td>c.823C&gt;T:p.Arg275Trp</td>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>PARK7</td>
<td>c.310G&gt;A:p.Ala104Thr</td>
<td>N</td>
<td>1</td>
</tr>
</tbody>
</table>

Abbreviations: CNV = copy number variant; PD = Parkinson disease; SNV = single nucleotide variant.

All indicated SNVs were heterozygous, except PRKN c.719C>T:p.Thr240Met, which was hemizygous, as the variant is in trans to a deletion allele. Pathogenic variants were considered diagnostic (Y) if discovered in an autosomal dominant gene, or in the case of autosomal recessive genes, if in combination with a CNV (asterisk, see also figure 3). Nondiagnostic (N), heterozygous SNVs were also discovered in PRKN (p.Arg275Trp) and PARK7 (p.Ala104Thr). In 2 subjects, SNVs in both LRRK2 and GBA were identified (double heterozygotes).

<sup>b</sup> LRRK2 p.Gly2019Ser and GBA p.Leu483Pro.
<sup>4</sup> Interpretations of c.1310C>T:p.Pro437Leu are conflicting (see e-References, links.lww.com/NXG/A306).
probands, increasing the overall genetic diagnostic yield to 19.3% (n = 21).

**Digital droplet PCR**

Given the observed high frequency of **PRKN** CNVs in this familial PD cohort (~5%), and the high cost of clinical-grade aCGH, we examined the feasibility of an exon-by-exon ddPCR assay to detect **PRKN** CNVs as a proof of principle. ddPCR is an emerging cost-effective method for sensitive and reliable assessment of specific CNVs. Indeed, ddPCR revealed all **PRKN** CNVs detected using aCGH (subjects 6, 11, 20, 21, and 22) (figure 4A). We also interrogated an additional 92 cases from our cohort with available DNA for intragenic **PRKN** CNVs using ddPCR, without discovery of additional CNVs (figure 4A and figure e-2, links.lww.com/NXG/A305). We also applied ddPCR to screen for potential CNVs at the **GBA** locus.

The 12-exon **GBA** shares high homology with the nearby 13-exon pseudogene, **GBAP1** (figure 4B). As shown in figure 4C, using ddPCR, we successfully amplified all 12 exons of **GBA**. Six exons (1, 2, 4, 7, 9, and 10) identified unique amplicons with a positive droplet ratio of 1, whereas the other 6 exons (3, 5, 6, 8, 11, and 12) demonstrated shared amplicons with the pseudogene, resulting in a droplet ratio of 2. Importantly, ddPCR also confirmed the deletion of **GBA** exons 2–8 in subject 1 (figure 4C) and did not reveal evidence for additional CNVs in 85 other samples tested (figure 4C and figure e-3, links.lww.com/NXG/A305). Of note, ddPCR initially suggested a single exon deletion (exon 6) in subject 48; however, further investigation using Sanger sequencing revealed an intronic SNV likely degrading ddPCR amplification (figure e-4, links.lww.com/NXG/A305). We redesigned the affected primer and demonstrated full amplification of exon 6 (e-Methods and figure e-4, links.lww.com/NXG/A305). Overall, our results suggest that ddPCR may be a sensitive and specific diagnostic tool for CNV detection in PD, including at loci such as **GBA** complicated by genomic regions with high sequence homology.

**CNV burden in clinical cohorts**

Compared with SNVs, limited reference data are available on the population frequency of CNVs, especially in...
neurologically healthy adult samples, hampering the interpretation of CNV frequencies detected in our cohort. We therefore leveraged data from the Baylor Genetics diagnostic laboratory, including 12,922 aCGH clinical referral samples. This large cohort is skewed for pediatric cases (mean age = 7.4 years, SD = 9.7 years, range 0–79 years), reflecting the more common use of aCGH in this population. Although the cohort includes a substantial proportion of individuals with developmental delay, autism, and dysmorphic features, there were no recorded submissions for PD. Based on stringent criteria (see e-Methods, links.lww.com/NXG/A305), at most PD loci, CNVs were either absent (DNAJC13, LRRK2, PINK1, and SNCA) or very rare at VPS35 (n = 2), GCH1 (n = 1), and PARK7 (n = 6, all subjects had 1p36 deletion syndrome). By contrast, CNVs were more common at 22q11.2 (n = 90, all losses a syndrome). By contrast, CNVs were more common at 22q11.2 (n = 90, all losses affecting the critical region) and PRKN (n = 95). Notably, the frequency of PRKN CNVs in our PD cohort (5.1%) represents a significant increase when compared with that of the Baylor Genetics clinical reference sample (frequency = 0.74%, OR = 7.2, 95% confidence interval 2.9–18.1, p = 2.8 × 10^{-5}). Because of the pseudogene, GBA1, and suboptimal probe coverage, the array does not reliably capture GBA CNVs.

**Discussion**

Establishing a specific genetic diagnosis can provide information about PD risk and progression relevant to patients and their families and may soon influence treatment decisions. In our familial PD sample, ES and aCGH independently identified a genetic cause for PD in 13.8% and 2.0%, respectively. The diagnostic yield for ES was slightly higher than that recently reported for an early-onset PD cohort (11.25%) and was also greater than the 10.7% diagnostic rate in an unselected adult series referred for clinical diagnostic ES. Given incipient treatment trials for GBA-PD and the potential importance of identifying eligible subjects in the future, our analyses considered lower-risk pathogenic alleles (OR ~2.4), p.Glu365Lys and p.Thr408Met, along with higher-penetrance variants (e.g., p.Leu483Pro, OR >5). Importantly, integrated ES and aCGH identified 5 additional subjects (4 unrelated probands)—including a subject with a GBA deletion—yielding an overall combined diagnostic rate of 19.3%. We also uncovered numerous VUS, including SNVs within Mendelian PD genes (table e-2, links. lww.com/NXG/A306) as well as large CNVs affecting other loci (table e-4, links.lww.com/NXG/A306). Although additional evidence will be required to confirm or refute pathogenicity, our genetic diagnostic rate would nearly double if these VUS in PD genes are bona fide risk factors. Overall, our findings suggest that integrated ES and aCGH analysis is essential for routine, high-confidence genetic diagnosis in familial PD.

Most genetic diagnostic studies in PD cohorts to date have ignored the potential contribution of CNVs. Similarly, except in several notable targeted CNV studies, research-based PD gene discovery has almost exclusively focused on SNVs, using ES or genotyping arrays. Importantly, we would have missed multiple pathogenic CNV alleles at both autosomal dominant (SNCA and GBA) and recessive (PRKN) loci without performing aCGH. In 5 subjects, pathogenic alleles discovered at PRKN would have been nondiagnostic based on isolated ES, leading to misclassification as heterozygous carriers, whereas integrated SNV-CNV analyses successfully established the molecular diagnosis of PRKN-PD. Our findings suggest caution for interpretation of studies attributing PD risk to either PRKN CNV or SNV heterozygous carrier states in isolation, consistent with prior studies. Although we did not detect any CNVs at PINK1 or PARK7 in our cohort, the importance of integrated SNV-CNV analysis may
extend to other autosomal recessive PD loci besides PRKN. Because our CNV and SNV data are unphased, and parental genotypes are not available, we cannot definitively exclude the possibility that certain CNVs and SNVs at PRKN were in cis- rather than trans-configuration. Nevertheless, our data suggest that structural variants may co-occur with SNVs more commonly than previously recognized, making consideration of both allele types important for comprehensive genetic diagnosis in PD.

ES has significantly accelerated the scope of gene discovery in PD and other neurologic disorders, but remains insensitive to allele classes such as trinucleotide repeat expansions and CNVs. Although bioinformatic tools may help identify CNVs from ES data, available algorithms have high false-positive rates when compared with aCGH, and this method may miss up to 30% of clinically relevant CNVs. To our knowledge, genome-wide aCGH with exon-by-exon coverage has not been previously applied in PD. Limitations of aCGH include significant cost and the possibility of missing small deletions/duplications. Alternative methods, such as ddPCR, may offer a cost-effective alternative for screening specific genes, including for small CNVs. In our study, ddPCR showed high sensitivity and specificity for detection of CNVs at both PRKN and GBA. Moreover, ddPCR successfully differentiated copy number changes affecting exons unique to

![aCGH plots and breakpoint junction sequences of PRKN CNVs](image-url)
GBA avoiding potential confounding by the adjacent pseudogene, GBAP1.

Despite evidence of an important role in disease risk, the mechanism(s) for generating CNVs relevant to PD remain largely unknown. Broadly, CNVs may form through mechanisms associated with DNA recombination, DNA replication, and/or DNA repair. Nonallelic homologous recombination can result in recurrent rearrangements. In contrast, nonhomologous end joining, fork stalling and template switching (FoSTeS) and microhomology-mediated break-induced replication (MMBIR), lead to nonrecurrent CNVs. In our study, all junction breakpoint sequencing results were consistent with the FoSTeS/MMBIR mechanism, including for CNV alleles discovered at SNCA, PRKN, and GBA. These results are consistent with a recent analysis of SNCA duplications from 6 independent cases. These findings have important implications for screening assays because the detection of nonrecurrent CNVs requires methods sensitive for heterogeneous, exon-by-exon changes. The FoSTeS/MMBIR mechanism can also trigger multiple iterative template switches in a single event, leading to the generation of more complex genomic rearrangements. Breakpoint sequencing of an SNCA CNV first observed in the Spellman-Muenter/Iowa

**Figure 4 PRKN and GBA ddPCR results of representative subjects**

(A) Positive droplet concentrations in 8 subjects. Primer pairs for the 12 exons of PRKN and 2 control genes, RPPH1 and TERT, were used to obtain positive droplet concentrations from PCR in each individual (e-Methods and figure e-4A, links.lww.com/NXG/A305). The y-axis shows exon-by-exon results in 13 columns with different colors, showing comparable results to the average value of RPPH1 and TERT. A y-axis value of 0.5 indicates a deletion, 1 copy neutral (no deletion, no duplication), and 1.5 a duplication. In subject 6, a duplication involving exons 4 to 6 was identified as shown by aCGH; in subject 11, exons 2, 4, 5, and 6 demonstrated copy number gains; in subject 20, there is a copy number loss involving exons 8 and 9; similarly, in subjects 21 and 22 a copy number loss of exons 5 and 6 is detected. In subjects 1, 23, and HapMap NA10851, no amplicons showed altered copy number. See also figure e-2 (links.lww.com/NXG/A305). Copy number variants are denoted with asterisks (*). (B) GBA and its nearby pseudogene, GBAP1, share a high degree of sequence homology, with ddPCR primer pairs for 6 of the 12 exons of GBA producing amplicons concurrently from GBA and GBAP1. GBA exons 3, 5, 6, 8, 11, and 12 are color coded to demonstrate their homologous regions within GBAP1, which result in a doubling of the apparent copy number identified by ddPCR: 4 instead of 2 copies (ratio = 2), indicate copy number neutrality for these exons. GBA exon 5 is homologous with an intragenic region between exons 4 and 5 of GBAP1. (C) ddPCR detected potential exonic CNVs in GBA. Here, we demonstrate a deletion identified in subject 1, compared with HapMap subject NA10581 and other 2 subjects, ratios of exons 2 to 8 were each reduced by 0.5-fold, consistent with a deletion involving these exons. Deleted exons are denoted with an asterisk (*); deleted exons with a droplet ratio of 1.5 due to GBAP1 amplification are denoted with an arrowhead. See also figure e-3 (links.lww.com/NXG/A305). aCGH = array-based comparative genomic hybridization; CNV = copy number variant; ddPCR = droplet digital PCR.
kindred confirmed the DUP-TRP-DUP structure and further revealed an internal inversion (DUP-TRP/INV-DUP) and microhomology. This rearrangement must have arisen during mitosis via FoSTeS/MMBIR, and therefore likely represents a de novo triplication, in contrast to the meiotic PMP22 triplications observed in Charcot-Marie-Tooth (MIM#118220), which derives from a duplication in the previous generation. Our results therefore demonstrate the essential role of breakpoint junction sequencing in definitively resolving CNV structure and responsible mechanisms.

Deletions in GBA rarely contribute to autosomal recessive Gaucher disease (MIM#230800). Our discovery of a GBA deletion allele in a subject with PD, expected to cause glucocerebrosidase haploinsufficiency, adds to other emerging evidence supporting a loss-of-function mechanism in GBA-PD. It will be informative to screen for additional GBA CNVs in additional case/control cohorts—perhaps using ddPCR—to determine how commonly these alleles are associated with PD risk and estimate their effect size and penetrance. We also identified 2 subjects doubly heterozygous for SNVs in both GBA and LRRK2, consistent with prior reports. We expect that additional PD cases compatible with oligogenic inheritance models will emerge following widespread adoption of comprehensive, genome-wide diagnostic approaches, including ES and aCGH. Future studies must address how such alleles may interact to modify PD risk and/or clinical manifestations. Finally, although our study focused on pathogenic alleles in established Mendelian loci, future assessment of a more complete spectrum of genetic variation through integrated SNV-CNV analysis is also likely to enhance power for novel PD gene discovery.

**Acknowledgment**

The authors thank all the families who participated in this study. They also thank Dr. Dennis W. Dickson and Dr. Zbigniew Wszolek (Mayo Clinic, Jacksonville, FL) for providing the SNCA-positive control sample. They also acknowledge Anita Kaw (Baylor College of Medicine, Houston, TX) for help with sample preparation and Mitchell A. Rao and Ido Machol (Baylor Genetics, Houston, TX) for data retrieval from the clinical database.

**Study funding**

This work was supported in part by The American Academy of Neurology (L.A.R.), Baylor College of Medicine Medical Genetics Research Fellowship T32 GM07526 (L.A.R.), NINDS K08 NS112467 (L.A.R.), the Baylor-Hopkins Center for Mendelian Genomics, BHCMG (UM1 HG006542 to J.R.L.), which is jointly funded by the National Human Genome Research Institute (NHGRI), National Heart Lung and Blood Institute (NHLBI), and NINDS R35 NS105078 (J.R.L.). J.E.P. was supported by NHGRI K08 HG008986. J.M.S. was supported by grants from the Huntington Foundation, Jan and Dan Duncan Neurological Research Institute at Texas Children’s Hospital, and a Career Award for Medical Scientists from the Burroughs Wellcome Fund. The Mayo Clinic Brain Bank is supported by the Mangurian Foundation Lewy Body Dementia Program at Mayo Clinic.

**Disclosure**


**Publication history**

Received by Neurology: Genetics December 27, 2019. Accepted in final form June 15, 2020.

**Appendix Authors**

<table>
<thead>
<tr>
<th>Name</th>
<th>Location</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laurie A. Robak, MD, PhD</td>
<td>Baylor College of Medicine, Houston, TX</td>
<td>Conceptualization, investigation, data generation, data analysis, data curation, writing—original draft, and writing—review and editing</td>
</tr>
<tr>
<td>Renqian Du, PhD, DDS</td>
<td>Baylor College of Medicine, Houston, TX</td>
<td>Conceptualization, investigation, data generation, data analysis, data curation, writing—original draft, and writing—review and editing</td>
</tr>
<tr>
<td>Bo Yuan, PhD</td>
<td>Baylor College of Medicine, Houston, TX</td>
<td>Conceptualization, investigation, data generation, data analysis, data curation, and writing—review and editing</td>
</tr>
<tr>
<td>Shen Gu, PhD</td>
<td>Baylor College of Medicine, Houston, TX</td>
<td>Conceptualization, investigation, data generation, data analysis, data curation, and writing—review and editing</td>
</tr>
<tr>
<td>Isabel Alfradique-Dunham, MD</td>
<td>Baylor College of Medicine, Houston, TX</td>
<td>Data curation and writing—review and editing</td>
</tr>
<tr>
<td>Vismaya Kondapalli, BS</td>
<td>Baylor College of Medicine, Houston, TX</td>
<td>Data generation, writing—original draft, and writing—review and editing</td>
</tr>
<tr>
<td>Evelyn Hinojosa, BS</td>
<td>Baylor College of Medicine, Houston, TX</td>
<td>Data curation and writing—review and editing</td>
</tr>
<tr>
<td>Amanda Stillwell, BS</td>
<td>Baylor College of Medicine, Houston, TX</td>
<td>Data curation and writing—review and editing</td>
</tr>
<tr>
<td>Emily Young, BA</td>
<td>Baylor College of Medicine, Houston, TX</td>
<td>Data curation and writing—review and editing</td>
</tr>
</tbody>
</table>

Continued
## References


Integrated sequencing and array comparative genomic hybridization in familial Parkinson disease
Laurie A. Robak, Renqian Du, Bo Yuan, et al.

Neurol Genet 2020;6;
DOI 10.1212/NXG.0000000000000498

This information is current as of July 28, 2020

Updated Information & Services
including high resolution figures, can be found at:
http://ng.neurology.org/content/6/5/e498.full.html

References
This article cites 37 articles, 1 of which you can access for free at:
http://ng.neurology.org/content/6/5/e498.full.html##ref-list-1

Permissions & Licensing
Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at:
http://ng.neurology.org/misc/about.xhtml#permissions

Reprints
Information about ordering reprints can be found online:
http://ng.neurology.org/misc/addir.xhtml#reprintsus