

Integrated Sequencing & Array Comparative Genomic Hybridization in Familial Parkinson's Disease

Laurie A. Robak, MD, PhD,^{1,*}, Renqian Du, PhD, DDS^{1,*}, Bo Yuan, PhD¹, Shen Gu, PhD¹, Isabel Alfradique-Dunham, MD², Vismaya Kondapalli, BS¹, Evelyn Hinojosa, BS¹, Amanda Stillwell, BS¹, Emily Young, BA¹, Chaofan Zhang, BS¹, Xiaofei Song, PhD¹, Haowei Du, MS¹, Tomasz Gambin, PhD¹, Shalini N. Jhangiani, PhD³, Zeynep Coban Akdemir, PhD^{1,4}, Donna M. Muzny, MS³, Anusha Tejomurtula, MS¹, Owen A. Ross, PhD^{5,6,7}, Chad Shaw, PhD¹, Joseph Jankovic, MD^{2,8}, Weimin Bi, PhD^{1,4}, Jennifer E. Posey MD, PhD¹, James R. Lupski, MD, PhD^{1,3,9,10,+ #}, Joshua M. Shulman MD, PhD^{1,2,11,12,+,#}

¹Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, 77030, USA

²Department of Neurology, Baylor College of Medicine, Houston, TX 77030, USA

³Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA

⁴Baylor Genetics, Houston, TX, 77021, USA

⁵Department of Neurology, Mayo Clinic, Jacksonville, FL, 32224, USA

⁶Department of Neuroscience, Mayo Clinic, Jacksonville, FL, 32224, USA

⁷Department of Clinical Genomics, Mayo Clinic, Jacksonville, FL, 32224, USA

⁸Parkinson's Disease Center and Movement Disorders Clinic, Baylor College of Medicine, Houston, TX 77030, USA

⁹Department of Pediatrics, Baylor College of Medicine, Houston, TX, 77030, USA

¹⁰Department of Pediatrics, Texas Children's Hospital, Houston, TX, 77030, USA

¹¹Department of Neuroscience, Baylor College of Medicine, Houston, TX, 77030, USA

¹²Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX, USA

*Contributed equally, co-first authors;

+Contributed equally, co-last authors

#Correspondence:

Joshua M. Shulman, MD, PhD
Jan and Dan Duncan Neurological Research Institute
1250 Moursund St., Suite N.1150
Houston, TX 77030
(832) 824-8976 / Joshua.Shulman@bcm.edu

James Lupski, MD, PhD, D.Sc. (hon)
Baylor College of Medicine
One Baylor Plaza, 6048
Houston, TX 77030
713-798-6530 / jlupski@bcm.edu

Running Title: Exome and aCGH in familial Parkinson's disease

Word Count: 3,734

Key Words: Parkinson's disease, exome sequencing, copy number variants

Financial Disclosure/ Potential Conflict of Interest:

JRL has stock ownership in 23andMe and Lasergen, is a paid consultant for Regeneron, and a co-inventor on multiple United States and European patents related to molecular diagnostics for inherited neuropathies, eye diseases and bacterial genomic fingerprinting. JMS consults for the Adrienne Helis Malvin & Diana Helis Henry Medical Research Foundations. Other authors declare that they have no conflict of interest.

Funding Sources:

This work was supported in part by The American Academy of Neurology (LAR), Baylor College of Medicine Medical Genetics Research Fellowship T32 GM07526 (LAR), the Baylor-Hopkins Center for Mendelian Genomics, BHCMG (UM1 HG006542, to JRL), which is jointly funded by the National Human Genome Research Institute (NHGRI), National Heart Lung and Blood Institute (NHLBI) and National Institute of Neurological Disorders and Stroke (NINDS) R35 NS105078 (to JRL). JEP was supported by NINDS R35 NS105078 and NHGRI K08 HG008986. JMS was supported by grants from Huffington 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.

Abstract

Background: Parkinson's disease (PD) is a genetically heterogeneous condition; both single nucleotide variants (SNVs) and copy number variants (CNVs) are important genetic risk factors.

We examined the utility of combining exome sequencing and genome-wide array-based comparative genomic hybridization (aCGH) for identification of PD genetic risk factors.

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

Results: Using exome sequencing, we discovered individuals with known pathogenic single nucleotide variants in *GBA* (p.E365K, p.T408M, p.N409S, p.L483P) and *LRRK2* (p.R1441G and p.G2019S). 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.N52fs, p.T240M, p.P437L, p.W453*) and likely disrupting CNVs at the *PARK2* 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 exome sequencing and aCGH yielded a genetic diagnosis in 19.3% of our familial PD cohort. Our analyses highlight potential mechanisms for *SNCA* and *PARK2* CNV formation, uncover multilocus pathogenic variation, and identify novel SNVs and CNVs for further investigation as potential PD risk alleles.

Introduction

Parkinson's disease (PD) is the second most common adult-onset neurodegenerative disorder with substantial evidence for heritability.¹⁻³ Up to 20% of patients with PD report a positive family history⁴⁻⁶ and genetic risk factors are more common in these families.⁷ More than 40 different loci that increase PD susceptibility have been identified in familial and sporadic PD⁸⁻¹². Although PD is diagnosed based on clinical criteria,¹³ identification of specific genetic risk factors can reveal prognostic information, such as risk of cognitive impairment and/or rate of progression,^{14,15} and may soon highlight eligibility for personalized therapies.^{16,17} In addition, discovery of risk variants may inform genetic counseling of unaffected family members. Indeed, surveys of patients and caregivers reveal a high level of interest in genetic testing for PD.^{18,19}

Exome sequencing (ES) has accelerated the discovery of PD genetic risk factors²⁰⁻²² and is ideally suited to identify single nucleotide variants (SNVs) in genetically heterogeneous diseases. In one study of adult patients referred for diverse clinical indications, ES had a diagnostic yield of 10% in individuals over 30 years of age.²³ 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%.²⁴ 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.²⁵ Despite notable recent advances,²⁶ ES remains insensitive for detection of small CNVs (< 50Kb).²⁷ Several complementary approaches, including multiplex ligation-dependent probe amplification,²⁸ bacterial artificial chromosome arrays,²⁹ and single nucleotide polymorphism arrays^{30,31} 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.^{32–35,34,35} While not yet adopted in most diagnostic laboratories, droplet digital Polymerase Chain Reaction (ddPCR) is also emerging as a rapid and cost-efficient, targeted approach for the assessment of small CNVs at specific loci.³⁶ Compared to standard quantitative PCR, digital PCR offers enhanced copy number and gene dosage sensitivity, precision, and reliability due to sample partitioning.³⁷

We explored the genetic molecular diagnostic rate for integrated ES and aCGH in a familial PD cohort. In addition, we evaluate ddPCR for confirmation of pathogenic CNVs, and using breakpoint sequencing, we elucidate potential mechanisms for CNV formation.

Methods

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 (45% of cases reported an affected first degree relative; 55% reported a second or third degree relative). All diagnoses were made by movement disorder specialists. Subjects in our cohort were unrelated, except for 2 brothers (subjects 21 and 22) known to have *PARK2*-related PD.¹² However, neither ES nor aCGH were previously performed in this sibling pair. Parental and other family member samples were not available for any subjects. All subjects provided informed consent and the study was approved by the BCM Institutional Review Board. As a positive control for aCGH, we also included a sample from a known subject with an *SNCA* triplication.^{38–41} We also interrogated a Baylor Genetics diagnostic laboratory sample including 12,922 clinical referral samples for

aCGH from peripheral blood.⁴² This analysis of aggregate clinical genomic data was also approved by the BCM Institutional Review Board. Subject numbers throughout the text are consistent with clinical and demographic details provided in Supplementary Table 1.

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* (MIM#168601),⁴³ *GBA* (MIM#168600),^{44,45} *LRRK2* (MIM#607060),^{46–48} *GCH1* (MIM#600225),²⁰ *DNAJC13* (MIM#616361),^{49,50} and *VPS35* (MIM#614203),^{51,52} and the autosomal recessive loci, *PARK2* (*PRKN*, MIM#600116),⁵³ *PINK1* (MIM#605909)⁵⁴ and *PARK7* (*DJI*, MIM#606324),⁵⁵ based on the available literature in April 2015 when this study was initiated. Deletions of 22q11.2 have recently been recognized as incompletely-penetrant PD risk factors^{56,57} and were also considered for CNV analyses. From the ES data, all candidate variants were reviewed by Clinical Geneticists (LR, JP, JL) and Movement Disorder Neurologists (JS, JJ) to establish a consensus on pathogenic alleles, integrating data from multiple available resources (PubMed, ClinVar⁵⁸, Human Genome Mutation Database⁵⁹). All pathogenic alleles included in this study are well-established, non-synonymous coding variants with moderate to high penetrance (OR>2) meeting stringent evidence for replication across studies or within the same study. All other variants discovered in these genes but not previously reported in PD were considered variants of unknown significance (VUSs).

SNV and CNV Detection

Genomic DNA was extracted from peripheral blood samples obtained from each participant. ES for the first 50 subjects was obtained from a prior study²¹, and ES for the remaining 60 subjects

was performed at the BCM Human Genome Sequencing Center (HGSC) using the Illumina HiSeq 2000.⁶⁰ Samples achieved an average of 95% of targeted exome bases covered to a depth of 20X or greater. All pathogenic SNVs detected via ES were confirmed via Sanger sequencing. Genome-wide array CGH was performed on 99 out of 110 subjects for which sufficient DNA remained after ES, using Baylor Genetics V10 2x400K high-density clinical-grade oligonucleotide microarrays, which were previously used for clinical diagnostic purposes. We defined potential CNVs as those regions with three or more consecutive probes with consistent direction of effect. To confirm these potential CNVs, we employed a custom 8x60K high-density array through Agilent (CA, USA). For the Baylor Genetics clinical cohort lookup, data was available from 12,922 subjects profiled for CNVs using Baylor arrays (v9 or v10).⁴² CNVs were filtered according to the following criteria: 3+ consecutive probes with consistent direction of effect, \log_2 ratios for loss ≤ 0.5 , gain > 0.35 . Filtered CNVs were manually assessed to exclude artifacts.⁶¹ For comparison of CNV frequencies, the 2-tailed chi square test was employed. See also Supplemental Methods for further details, including ddPCR protocol.

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.G2019S and c.4321C>G:p.R1441G) and *GBA* (c.1093G>A:p.E365K, c.1223C>T:p.T408M, c.1448T>C:p.L483P, and c.1226A>G:p.N409S) (Table 1). Two subjects each harbored heterozygous SNVs in both *GBA* and *LRRK2*; i.e. double heterozygotes. One subject had a combination of *LRRK2* p.G2019S and *GBA* p.E365K (Subject 2), while the other had *LRRK2* p.G2019S and *GBA* p.L483P (Subject 13). Both subjects had onset of PD symptoms in their 40s. On initial exam, 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 (Supplementary Figure 1A; Supplementary Table 1). 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 (Supplementary Figure 1B, Supplementary Table 1). 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 seven individuals with pathogenic variants in loci usually associated with autosomal recessive PD, including *PARK2* (6 individuals) and *DJ-1/PARK1* (1 individual). However, all subjects were heterozygous SNV carriers, and therefore isolated ES was non-diagnostic (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. All implicated variants were confirmed via Sanger sequencing. Besides the pathogenic variants noted above, ES also identified heterozygous variants of unknown significance (VUS) in many PD risk genes (Supplementary Table 2), including *LRRK2* (n=12), *GCH1* (n=1), *VPS35* (n=1), *DNAJC13* (n=18), *DJ-1/PARK7* (n=2),

PARK2 (n=9), and *PINK1* (n=1). For recessive PD genes, all subjects harbored a single heterozygous variant.

Copy Number Variants

We next interrogated aCGH data for pathogenic CNVs among PD genes. Our analyses included 99 out of 110 total subjects evaluated by ES. No CNVs were observed in *VPS35*, *LRRK2*, *DNAJC13*, *GCH1*, *DJ1*, or *PINK1*, nor did we identify any candidate deletions at the 22q11.2 locus. However, we discovered CNVs in *SNCA* (n=1), and *GBA* (n=1), and *PARK2* (n=5). All reported CNVs were confirmed by custom high-density arrays and breakpoint sequencing. We first discuss CNVs in *SNCA* and *GBA* since these affect dominant PD genes and were considered diagnostic based on aCGH alone. All heterozygous CNVs in the recessive PD gene, *PARK2*, were considered further in combination with ES results (see next section). Overall, isolated aCGH identified a genetic risk factor for PD that was diagnostic in 2.0% of our cohort (n=2 of 99 probands). Based on aCGH, we have also detected numerous large CNVs (> 1 Mb) within our cohort that impact other genomic loci; these variants remain of uncertain clinical significance but are documented for completeness (Supplementary Table 3).

In Subject 3, we detected a 248 kb duplication encompassing *SNCA*, as well as the adjacent *MMRNI*, which has no known association with human disease. This CNV was confirmed by high-density aCGH and breakpoint analysis (Figure 1A). On initial exam, this subject exhibited rigidity, tremor, 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 two copies to one during the template switch accompanying replicative repair.^{62,63}

Microhomology has been reported as a signature for a class of DNA replication errors that can generate CNVs (see Discussion).^{64,65} 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.³⁸⁻⁴¹ 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 variant⁶⁶ and is also consistent with a likely replication-based mechanism for CNV formation.⁶⁷

We also discovered a heterozygous intragenic deletion in *GBA* in one subject (Figure 2), who presented at age 28 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 (Supplementary Figure 1C). Since variant confirmation at the *GBA* locus can be complicated by an adjacent pseudogene with significant homology,⁶⁸ we confirmed the 4.7 kb deletion of exons 2-8 using long-range PCR (Figure 2C). Breakpoint sequencing also confirmed heterozygosity and further revealed a 5-bp microhomology domain consistent with CNV formation due to non-homologous recombination or replication errors (Figure 2B).

Integrated Analysis of SNVs and CNVs

As highlighted above, isolated ES and aCGH each discovered a number of subjects with heterozygous SNVs or CNVs affecting recessive genes. In order to determine if these changes

might be diagnostic, we examined the results together for potential biallelic variation at the locus due to both a SNV allele and a CNV allele. Indeed, three subjects in our cohort were newly identified as potential compound heterozygous carriers of both a pathogenic CNV and SNV in *PARK2* (Figure 3). Subject 6 had a 364 kb duplication of exons 4-6 and a frameshift deletion c.155delA:p.N52fs. Subject 20 had a 222 kb deletion of exons 8-9 and a stopgain c.1358G>A:p.W453*. Subject 11 harbored a pathogenic *PARK2* SNV (c.1310C>T:p.P437L) and a complex locus rearrangement, including a copy number neutral region flanked by 404 kb and 199 kb duplications affecting exons 2-3 and exons 5-6, respectively (Figure 3B). Our cohort also included 2 brothers with known *PARK2* PD;¹² however, neither ES nor aCGH were previously performed. Our analysis confirmed compound heterozygosity for the known pathogenic SNV in exon 6 (c.719C>T:p.T240M, apparently homozygous on ES) and a CNV (178-kb deletion of exons 5-6) (Figure 3C). Interestingly, ES also discovered an additional VUS (c.2T>C:p.M1T). Based on available clinical information (Supplementary Table 1), all subjects with *PARK2* variants had young-onset PD (age range 15-36).

All CNVs were again confirmed 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 *PARK2* CNVs, junction structure is detailed in Figure 3, including implicated mechanisms of CNV formation. Overall, integrated ES and CNV identified a genetic cause for PD in 4 additional probands, increasing the overall genetic diagnostic yield to 19.3% (n=21).

Digital Droplet PCR (ddPCR)

Given the observed high frequency of *PARK2* CNVs in this familial PD cohort (~5%), and the high cost of clinical-grade aCGH, we examined the feasibility of exon-by-exon ddPCR assay to detect *PARK2* CNVs as a proof of principle. ddPCR is an emerging cost-effective method for sensitive, and reliable assessment of specific CNVs^{36,37}. Indeed, all *PARK2* CNVs described (subjects 6, 11, 20, 21 and 22) were robustly detected using ddPCR (Figure 4A). An additional 94 cases from our cohort with available DNA were interrogated for intragenic *PARK2* CNVs, and no additional CNVs were detected (Figure 4A and Supplementary Figure 2). We also applied ddPCR to screen for potential CNVs at the *GBA* locus. The 12-exon *GBA* shares high homology with a nearby 13-exon pseudogene, *GBAP1*. 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, while the other six exons (3, 5, 6, 8, 11, and 12) demonstrated shared amplicons with the pseudogene, resulting in a droplet ratio of 2. Importantly, the deletion of *GBA* exons 2-8 in Subject 1 was successfully detected and confirmed by ddPCR (Figure 4C), and the assay was negative for *GBA* CNVs in 86 other samples tested (Figures 4C and Supplementary Figure 3). Of note, a single exon deletion (exon 6) was suggested by ddPCR in subject 48. Further investigation revealed an intronic SNV located at the 3' end of one ddPCR primer affecting amplification efficacy, confirmed by Sanger sequencing (Supplementary Figure 4A). We re-designed the affected primer (Supplementary Methods) and demonstrated full amplification of exon 6 (Supplementary Figure 4B). Overall, our results suggest that ddPCR may be a sensitive and specific diagnostic tool for CNV detection in PD subjects, including at loci such as *GBA* complicated by genomic regions with high sequence homology.

CNV Burden in Clinical Cohorts

Compared to SNVs, limited reference data is 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 Methods), at most PD loci, CNVs were either absent (*DNAJC13*, *LRRK2*, *PINK1*, and *SNCA*) or very rare at *VPS35* (n=2), *GCH1* (n=1), and *DJ-1* (n=6, all subjects had 1p36 deletion syndrome). By contrast, CNVs were more common at 22q11.2 (n=90, all losses affecting the critical region) and *PARK2* (n=95). Notably, the frequency of *PARK2* CNVs in our PD cohort (5.1%) represents a significant enrichment ($p=8.5 \times 10^{-7}$), when compared to that of the Baylor Genetics clinical reference sample (0.74%). Due to the pseudogene, *GBAP1*, 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.^{17,69} We have evaluated combined ES and aCGH for genetic diagnosis in 110 subjects with familial PD (both assays were available for 99 subjects). In our cohort, ES and aCGH independently identified a genetic cause for PD in 13.8% and 2.0%, respectively. Our diagnostic yield for ES was slightly

higher than that recently reported for an early-onset PD cohort (11.25%),²⁴ and was also greater than the 10.7% diagnostic rate in an unselected adult series referred for clinical diagnostic ES.²³ Given incipient treatment trials for *GBA*-PD and the potential importance of identifying eligible subjects in the future⁶⁹, we included lower-risk pathogenic alleles (OR~2.4)⁷⁰, p.E365K and p.T408M, along with higher-penetrance variants (e.g. p.L483P, OR>5)⁴⁴. 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%. Our analyses also uncovered numerous VUS, including SNVs within Mendelian PD genes (Supplementary Table 2) as well as large CNVs affecting other loci (Supplementary Table 4). While additional evidence will be required to confirm or refute pathogenicity, our genetic diagnostic rate would nearly double if the variants in PD genes are considered *bona fide* risk factors. Overall, our findings suggest that complementary assessments for SNV and CNVs will be 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,^{29,30,56,71} research-based PD gene discovery has almost exclusively focused on SNVs, using ES or genotyping arrays.^{22,72} In our familial PD cohort, pathogenic alleles at both autosomal dominant (*SNCA*, *GBA*) and recessive (*PARK2*) loci would have been missed if not for the inclusion of aCGH for detection of CNVs. For example, out of 7 pathogenic alleles discovered using ES at recessive loci (*PARK2* and *DJ-1*), nearly all cases (except subjects 21 and 22) would have been non-diagnostic leading to misclassification as heterozygous carriers without aCGH. However, integrated SNV-CNV analyses successfully resolved 5 individuals with likely *PARK2* PD. Our findings suggest caution for interpretation of studies attributing PD risk to either *PARK2* CNV or

SNV heterozygous carrier states in isolation, consistent with prior studies.^{73,74} The importance of integrated SNV-CNV analysis likely extends to other autosomal recessive PD loci, including *DJ-1* and *PINK1*.^{55,75} Since 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 *PARK2* 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.⁷⁶

ES has significantly accelerated the scope of gene discovery in PD and other neurologic disorders,^{22,24} but remains insensitive to allele classes such as trinucleotide repeat expansions and CNVs. While bioinformatic tools can be employed to screen for CNVs in data from ES, available algorithms have high false positive rates compared to aCGH⁷⁷ and as many as 30% of clinically relevant CNVs are missed by ES.²⁷ 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,³⁶ including for small CNVs. In our study, ddPCR showed high sensitivity and specificity for detection of CNVs at both *PARK2* and *GBA*. In the latter case, ddPCR was able to robustly differentiate copy number changes affecting exons unique to *GBA* avoiding potential confounding by the adjacent pseudogene, *GBAP1*.

Despite evidence of an important role in disease risk, the mechanism(s) responsible for generating CNVs relevant to PD have been poorly studied. Broadly, CNVs may form through mechanisms associated with DNA recombination, DNA replication, and/or DNA repair.⁶² Mechanisms such as nonallelic homologous recombination^{78–80} can result in recurrent

rearrangements. In contrast, non-homologous end joining,^{64,81} fork stalling and template switching (FoSTeS) and microhomology-mediated break-induced replication (MMBIR),⁶⁴ lead to non-recurrent CNVs. In our study, for all CNVs for which breakpoint junctions were obtained, the mechanism was consistent with FoSTeS/MMBIR. This finding has important implications for screening assays since methods sensitive for heterogeneous, exon-by-exon changes must be employed to detect non-recurrent CNVs. 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 kindred,^{40,82} confirmed the DUP-TRP-DUP structure⁶⁶ and further revealed an internal inversion (DUP-TRP/INV-DUP) and also revealed microhomology. This rearrangement must have arisen during mitosis via FoSTeS/MMBIR^{62,64,83,84} 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.⁸⁵ Our results therefore demonstrate the essential role of breakpoint junction sequencing in definitively resolving CNV structure and responsible mechanisms.

To our knowledge, our study includes the first report of an intragenic *GBA* deletion allele in a subject with PD. Similar deletions in *GBA* have been rarely described in autosomal recessive Gaucher disease (MIM#230800).^{86–88} Our discovery of a *GBA* deletion allele, expected to cause glucocerebrosidase haploinsufficiency, adds to other emerging evidence supporting a loss-of-function mechanism in *GBA*-associated PD.^{89,90} 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 in our cohort doubly heterozygous for SNVs in both *GBA* and

LRRK2, and similar PD cases have been previously reported.⁹¹ As comprehensive, genome-wide diagnostic approaches, including ES and aCGH, become widely utilized in PD, it is likely that additional cases compatible with oligogenic inheritance models will be recognized.^{92,93} Future studies will be required to define whether and how such alleles may interact to modify PD risk and/or clinical manifestations. Finally, while 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.

Authors' Roles

1. Research project: A. Conception: LAR, JRL, JMS B. Organization: LR, JRL, JEP, JMS C. Execution; LAR, RD, BY, SG, IAD, VK, EH, AS, EY, CZ, XS, HD, TG, SNJ, ZCA, DMM, AT, OAR, CS, JJ, WB, JRL, JMS
2. Statistical Analysis: A. Design: LR, WB, JEP B. Execution: LR, WB, CS, JEP C. Review and Critique:: LAR, RD, BY, SG, IAD, VK, EH, AS, EY, CZ, XS, HD, TG, SNJ, ZCA, DMM, AT, OAR, CS, JJ, WB, JRL, JEP, JMS
3. Manuscript: A. Writing of the first draft: LAR, RD, VK, JEP, JRL, JMS. Review and Critique: LAR, RD, BY, SG, IAD, VK, EH, AS, EY, CZ, XS, HD, TG, SNJ, ZCA, DMM, AT, OAR, CS, JJ, WB, JRL, JEP, JMS

Financial Disclosures of All Authors

JRL has stock ownership in 23andMe and Lasergen, is a paid consultant for Regeneron, and a co-inventor on multiple United States and European patents related to molecular diagnostics for

inherited neuropathies, eye diseases and bacterial genomic fingerprinting. JMS consults for the Adrienne Helis Malvin & Diana Helis Henry Medical Research Foundations. Other authors declare that they have no conflict of interest.

Acknowledgements

We thank all of the families who participated in this study. We also thank Drs. Dennis W. Dickson and Zbigniew Wszolek for providing the *SNCA* positive control sample. We acknowledge Mitchell A. Rao and Ido Machol for data retrieval from the clinical database.

Supplementary Material

Supplementary material includes methods, 3 tables, and 4 figures.

References

1. Kalia L V, Lang AE. Parkinson disease in 2015: Evolving basic, pathological and clinical concepts in PD. *Nat Rev Neurol*. 2016;12(2):65-66. doi:10.1038/nrneurol.2015.249
2. Lesage S, Brice A. Parkinson's disease: from monogenic forms to genetic susceptibility factors. *Hum Mol Genet*. 2009;18(R1):R48-R59. doi:10.1093/hmg/ddp012
3. Obeso JA, Stamelou M, Goetz CG, et al. Past, present, and future of Parkinson's disease: A special essay on the 200th Anniversary of the Shaking Palsy. *Mov Disord*. 2017;32(9):1264-1310. doi:10.1002/mds.27115
4. Elbaz A, Grigoletto F, Baldereschi M, et al. Familial aggregation of Parkinson's disease: a population-based case-control study in Europe. EUROPARKINSON Study Group. *Neurology*. 1999;52(9):1876-1882. <http://www.ncbi.nlm.nih.gov/pubmed/10371537>.

Accessed March 19, 2019.

5. Kurz M, Alves G, Aarsland D, Larsen JP. Familial Parkinson's disease: a community-based study. *Eur J Neurol*. 2003;10(2):159-163.
<http://www.ncbi.nlm.nih.gov/pubmed/12603291>. Accessed March 20, 2019.
6. McDonnell SK, Schaid DJ, Elbaz A, et al. Complex segregation analysis of Parkinson's disease: The Mayo Clinic Family Study. *Ann Neurol*. 2006;59(5):788-795.
doi:10.1002/ana.20844
7. Karimi-Moghadam A, Charsouei S, Bell B, Jabalameli MR. Parkinson Disease from Mendelian Forms to Genetic Susceptibility: New Molecular Insights into the Neurodegeneration Process. *Cell Mol Neurobiol*. 2018;38(6):1153-1178.
doi:10.1007/s10571-018-0587-4
8. Brás J, Guerreiro R, Hardy J. SnapShot: Genetics of Parkinson's disease. *Cell*. 2015;160(3):570-570.e1. doi:10.1016/j.cell.2015.01.019
9. Verstraeten A, Theuns J, Van Broeckhoven C. Progress in unraveling the genetic etiology of Parkinson disease in a genomic era. *Trends Genet*. 2015;31(3):140-149.
doi:10.1016/j.tig.2015.01.004
10. Hernandez DG, Reed X, Singleton AB. Genetics in Parkinson disease: Mendelian versus non-Mendelian inheritance. *J Neurochem*. 2016;139 Suppl 1:59-74.
doi:10.1111/jnc.13593
11. Singleton AB, Farrer MJ, Bonifati V. The genetics of Parkinson's disease: Progress and therapeutic implications. *Mov Disord*. 2013;28(1):14-23. doi:10.1002/mds.25249
12. Deng H, Le W-D, Hunter CB, et al. Heterogeneous phenotype in a family with compound heterozygous parkin gene mutations. *Arch Neurol*. 2006;63(2):273-277.

doi:10.1001/archneur.63.2.273

13. Jankovic J. Parkinson's disease: clinical features and diagnosis. *J Neurol Neurosurg Psychiatry*. 2008;79(4):368-376. doi:10.1136/jnnp.2007.131045
14. Liu G, Boot B, Locascio JJ, et al. Specifically neuropathic Gaucher's mutations accelerate cognitive decline in Parkinson's. *Ann Neurol*. 2016;80(5):674-685. doi:10.1002/ana.24781
15. Saunders-Pullman R, Mirelman A, Alcalay RN, et al. Progression in the *LRRK2* - Associated Parkinson Disease Population. *JAMA Neurol*. 2018;75(3):312. doi:10.1001/jamaneurol.2017.4019
16. Jankovic J. Pathogenesis-targeted therapeutic strategies in Parkinson's disease. *Mov Disord*. 2019;34(1):41-44. doi:10.1002/mds.27534
17. Sardi SP, Cedarbaum JM, Brundin P. Targeted Therapies for Parkinson's Disease: From Genetics to the Clinic. *Mov Disord*. 2018;33(5):684-696. doi:10.1002/mds.27414
18. Scuffham TM, McInerney-Leo A, Ng S-K, Mellick G. Knowledge and attitudes towards genetic testing in those affected with Parkinson's disease. *J Community Genet*. 2014;5(2):167-177. doi:10.1007/s12687-013-0168-7
19. Falcone DC, Wood EM, Xie SX, Siderowf A, Van Deerlin VM. Genetic Testing and Parkinson Disease: Assessment of Patient Knowledge, Attitudes, and Interest. *J Genet Couns*. 2011;20(4):384-395. doi:10.1007/s10897-011-9362-0
20. Mencacci NE, Isaias IU, Reich MM, et al. Parkinson's disease in GTP cyclohydrolase 1 mutation carriers. *Brain*. 2014;137(9):2480-2492. doi:10.1093/brain/awu179
21. Farlow JL, Robak LA, Hetrick K, et al. Whole-Exome Sequencing in Familial Parkinson Disease. *JAMA Neurol*. 2016;73(1):68. doi:10.1001/jamaneurol.2015.3266
22. Jansen IE, Ye H, Heetveld S, et al. Discovery and functional prioritization of Parkinson's

- disease candidate genes from large-scale whole exome sequencing. *Genome Biol.* 2017;18(1):22. doi:10.1186/s13059-017-1147-9
23. Posey JE, Rosenfeld JA, James RA, et al. Molecular diagnostic experience of whole-exome sequencing in adult patients. *Genet Med.* 2016;18(7):678-685. doi:10.1038/gim.2015.142
24. Schormair B, Kemlink D, Mollenhauer B, et al. Diagnostic exome sequencing in early-onset Parkinson's disease confirms *VPS13C* as a rare cause of autosomal-recessive Parkinson's disease. *Clin Genet.* 2018;93(3):603-612. doi:10.1111/cge.13124
25. 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. doi:10.1007/s00439-016-1749-4
26. Gambin T, Akdemir ZC, Yuan B, et al. Homozygous and hemizygous CNV detection from exome sequencing data in a Mendelian disease cohort. *Nucleic Acids Res.* 2017;45(4):1633-1648. doi:10.1093/nar/gkw1237
27. Dharmadhikari A V, Ghosh R, Yuan B, et al. Copy number variant and runs of homozygosity detection by microarrays enabled more precise molecular diagnoses in 11,020 clinical exome cases. *Genome Med.* 2019;11(1):30. doi:10.1186/s13073-019-0639-5
28. Scarciolla O, Brancati F, Valente EM, et al. Multiplex ligation-dependent probe amplification assay for simultaneous detection of Parkinson's disease gene rearrangements. *Mov Disord.* 2007;22(15):2274-2278. doi:10.1002/mds.21532
29. Kim J-S, Yoo J-Y, Lee K-S, et al. Comparative Genome Hybridization Array Analysis for Sporadic Parkinson's Disease. *Int J Neurosci.* 2008;118(9):1331-1345.

doi:10.1080/00207450802174522

30. Pankratz N, Dumitriu A, Hetrick KN, et al. Copy Number Variation in Familial Parkinson Disease. Roberts RG, ed. *PLoS One*. 2011;6(8):e20988.

doi:10.1371/journal.pone.0020988

31. Simon-Sanchez J, Scholz S, Matarin M del M, et al. Genomewide SNP assay reveals mutations underlying Parkinson disease. *Hum Mutat*. 2008;29(2):315-322.

doi:10.1002/humu.20626

32. Wiszniewska J, Bi W, Shaw C, et al. Combined array CGH plus SNP genome analyses in a single assay for optimized clinical testing. *Eur J Hum Genet*. 2014;22(1):79-87.

doi:10.1038/ejhg.2013.77

33. Cheung SW, Shaw CA, Yu W, et al. Development and validation of a CGH microarray for clinical cytogenetic diagnosis. *Genet Med*. 2005;7(6):422-432.

doi:10.109701.GIM.0000170992.63691.32

34. Miller DT, Adam MP, Aradhya S, et al. Consensus Statement: Chromosomal Microarray Is a First-Tier Clinical Diagnostic Test for Individuals with Developmental Disabilities or Congenital Anomalies. *Am J Hum Genet*. 2010;86(5):749-764.

doi:10.1016/j.ajhg.2010.04.006

35. Harel T, Lupski JR. Genomic disorders 20 years on-mechanisms for clinical manifestations. *Clin Genet*. 2018;93(3):439-449. doi:10.1111/cge.13146

36. Gu S, Posey JE, Yuan B, et al. Mechanisms for the Generation of Two Quadruplications Associated with Split-Hand Malformation. *Hum Mutat*. 2016;37(2):160-164.

doi:10.1002/humu.22929

37. Härmälä SK, Butcher R, Roberts CH. Copy Number Variation Analysis by Droplet

- Digital PCR. In: *Methods in Molecular Biology (Clifton, N.J.)*. Vol 1654. ; 2017:135-149.
doi:10.1007/978-1-4939-7231-9_9
38. Book A, Guella I, Candido T, et al. A Meta-Analysis of α -Synuclein Multiplication in Familial Parkinsonism. *Front Neurol*. 2018;9:1021. doi:10.3389/fneur.2018.01021
39. Singleton AB, Farrer M, Johnson J, et al. -Synuclein Locus Triplication Causes Parkinson's Disease. *Science (80-)*. 2003;302(5646):841-841.
doi:10.1126/science.1090278
40. Spellman GG. Report of familial cases of parkinsonism. Evidence of a dominant trait in a patient's family. *JAMA*. 1962;179:372-374.
<http://www.ncbi.nlm.nih.gov/pubmed/13915636>. Accessed June 28, 2019.
41. Muentert MD, Forno LS, Hornykiewicz O, et al. Hereditary form of parkinsonism--dementia. *Ann Neurol*. 1998;43(6):768-781. doi:10.1002/ana.410430612
42. Gambin T, Yuan B, Bi W, et al. Identification of novel candidate disease genes from de novo exonic copy number variants. *Genome Med*. 2017;9(1):83. doi:10.1186/s13073-017-0472-7
43. 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.
<http://www.ncbi.nlm.nih.gov/pubmed/9197268>. Accessed March 20, 2019.
44. Sidransky E, Nalls MA, Aasly JO, et al. Multicenter Analysis of Glucocerebrosidase Mutations in Parkinson's Disease. *N Engl J Med*. 2009;361(17):1651-1661.
doi:10.1056/NEJMoa0901281
45. Nalls MA, Duran R, Lopez G, et al. A Multicenter Study of Glucocerebrosidase Mutations in Dementia With Lewy Bodies. *JAMA Neurol*. 2013;70(6):727.

doi:10.1001/jamaneurol.2013.1925

46. Funayama M, Hasegawa K, Kowa H, Saito M, Tsuji S, Obata F. A new locus for Parkinson's disease (PARK8) maps to chromosome 12p11.2-q13.1. *Ann Neurol*. 2002;51(3):296-301. <http://www.ncbi.nlm.nih.gov/pubmed/11891824>. Accessed March 19, 2019.
47. Paisán-Ruiz C, Jain S, Evans EW, et al. Cloning of the Gene Containing Mutations that Cause PARK8-Linked Parkinson's Disease. *Neuron*. 2004;44(4):595-600.
doi:10.1016/j.neuron.2004.10.023
48. Zimprich A, Biskup S, Leitner P, et al. Mutations in LRRK2 Cause Autosomal-Dominant Parkinsonism with Pleomorphic Pathology. *Neuron*. 2004;44(4):601-607.
doi:10.1016/j.neuron.2004.11.005
49. Vilariño-Güell C, Rajput A, Milnerwood AJ, et al. DNAJC13 mutations in Parkinson disease. *Hum Mol Genet*. 2014;23(7):1794-1801. doi:10.1093/hmg/ddt570
50. Gustavsson EK, Trinh J, Guella I, et al. DNAJC13 genetic variants in parkinsonism. *Mov Disord*. 2015;30(2):273-278. doi:10.1002/mds.26064
51. Vilariño-Güell C, Wider C, Ross OA, et al. VPS35 Mutations in Parkinson Disease. *Am J Hum Genet*. 2011;89(1):162-167. doi:10.1016/j.ajhg.2011.06.001
52. Zimprich A, Benet-Pagès A, Struhal W, et al. A mutation in VPS35, encoding a subunit of the retromer complex, causes late-onset Parkinson disease. *Am J Hum Genet*. 2011;89(1):168-175. doi:10.1016/j.ajhg.2011.06.008
53. Kitada T, Asakawa S, Hattori N, et al. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature*. 1998;392(6676):605-608. doi:10.1038/33416
54. Valente EM, Abou-Sleiman PM, Caputo V, et al. Hereditary Early-Onset Parkinson's

- Disease Caused by Mutations in PINK1. *Science* (80-). 2004;304(5674):1158-1160.
doi:10.1126/science.1096284
55. Bonifati V, Rizzu P, van Baren MJ, et al. Mutations in the DJ-1 Gene Associated with Autosomal Recessive Early-Onset Parkinsonism. *Science* (80-). 2003;299(5604):256-259.
doi:10.1126/science.1077209
56. Mok KY, Sheerin U, Simón-Sánchez J, et al. Deletions at 22q11.2 in idiopathic Parkinson's disease: a combined analysis of genome-wide association data. *Lancet Neurol*. 2016;15(6):585-596. doi:10.1016/S1474-4422(16)00071-5
57. Butcher NJ, Kiehl T-R, Hazrati L-N, et al. Association Between Early-Onset Parkinson Disease and 22q11.2 Deletion Syndrome. *JAMA Neurol*. 2013;70(11):1359.
doi:10.1001/jamaneurol.2013.3646
58. Landrum MJ, Lee JM, Riley GR, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*. 2014;42(D1):D980-D985.
doi:10.1093/nar/gkt1113
59. Stenson PD, Mort M, Ball E V, et al. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genet*. 2017;136(6):665-677.
doi:10.1007/s00439-017-1779-6
60. Posey JE, O'Donnell-Luria AH, Chong JX, et al. Insights into genetics, human biology and disease gleaned from family based genomic studies. *Genet Med*. January 2019;doi: 10.1038/s41436-018-0408-7. doi:10.1038/s41436-018-0408-7
61. Innes J, Reali L, Clayton-Smith J, et al. CNVs affecting cancer predisposing genes (CPGs) detected as incidental findings in routine germline diagnostic chromosomal microarray

- (CMA) testing. *J Med Genet.* 2018;55(2):89-96. doi:10.1136/jmedgenet-2017-104892
62. Carvalho CMB, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. *Nat Rev Genet.* 2016;17(4):224-238. doi:10.1038/nrg.2015.25
 63. Lupski JR. 2018 Victor A. McKusick Leadership Award: Molecular Mechanisms for Genomic and Chromosomal Rearrangements. *Am J Hum Genet.* 2019;104(3):391-406. doi:10.1016/j.ajhg.2018.12.018
 64. Zhang F, Khajavi M, Connolly AM, Towne CF, Batish SD, Lupski JR. The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. *Nat Genet.* 2009;41(7):849-853. doi:10.1038/ng.399
 65. Beck CR, Carvalho CMB, Akdemir ZC, et al. Megabase Length Hypermutation Accompanies Human Structural Variation at 17p11.2. *Cell.* 2019;176(6):1310-1324.e10. doi:10.1016/j.cell.2019.01.045
 66. 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 Park Dis.* 2018;4(1):18. doi:10.1038/s41531-018-0054-4
 67. Carvalho CMB, Ramocki MB, Pehlivan D, et al. Inverted genomic segments and complex triplication rearrangements are mediated by inverted repeats in the human genome. *Nat Genet.* 2011;43(11):1074-1081. doi:10.1038/ng.944
 68. Tsuji S, Choudary P V., Martin BM, et al. A Mutation in the Human Glucocerebrosidase Gene in Neuronopathic Gaucher's Disease. *N Engl J Med.* 1987;316(10):570-575. doi:10.1056/NEJM198703053161002
 69. Espay AJ, Brundin P, Lang AE. Precision medicine for disease modification in Parkinson disease. *Nat Rev Neurol.* 2017;13(2):119-126. doi:10.1038/nrneurol.2016.196

70. Benitez BA, Davis AA, Jin SC, et al. Resequencing analysis of five Mendelian genes and the top genes from genome-wide association studies in Parkinson's Disease. *Mol Neurodegener.* 2016;11(1):29. doi:10.1186/s13024-016-0097-0
71. La Cognata V, Morello G, Gentile G, et al. A customized high-resolution array-comparative genomic hybridization to explore copy number variations in Parkinson's disease. *Neurogenetics.* 2016;17(4):233-244. doi:10.1007/s10048-016-0494-0
72. Blauwendraat C, Faghri F, Pihlstrom L, et al. NeuroChip, an updated version of the NeuroX genotyping platform to rapidly screen for variants associated with neurological diseases. *Neurobiol Aging.* 2017;57:247.e9-247.e13. doi:10.1016/j.neurobiolaging.2017.05.009
73. Kay DM, Stevens CF, Hamza TH, et al. A comprehensive analysis of deletions, multiplications, and copy number variations in PARK2. *Neurology.* 2010;75(13):1189-1194. doi:10.1212/WNL.0b013e3181f4d832
74. Pankratz N, Kissell DK, Pauciulo MW, et al. Parkin dosage mutations have greater pathogenicity in familial PD than simple sequence mutations. *Neurology.* 2009;73(4):279-286. doi:10.1212/WNL.0b013e3181af7a33
75. Marongiu R, Brancati F, Antonini A, et al. Whole gene deletion and splicing mutations expand thePINK1 genotypic spectrum. *Hum Mutat.* 2007;28(1):98-98. doi:10.1002/humu.9472
76. Boone PM, Campbell IM, Baggett BC, et al. Deletions of recessive disease genes: CNV contribution to carrier states and disease-causing alleles. *Genome Res.* 2013;23(9):1383-1394. doi:10.1101/gr.156075.113
77. de Ligt J, Boone PM, Pfundt R, et al. Detection of Clinically Relevant Copy Number

- Variants with Whole-Exome Sequencing. *Hum Mutat.* 2013;34(10):1439-1448.
doi:10.1002/humu.22387
78. Stankiewicz P, Lupski JR. Genome architecture, rearrangements and genomic disorders. *Trends Genet.* 2002;18(2):74-82. <http://www.ncbi.nlm.nih.gov/pubmed/11818139>. Accessed March 20, 2019.
 79. Gu W, Zhang F, Lupski JR. Mechanisms for human genomic rearrangements. *Pathogenetics.* 2008;1(1):4. doi:10.1186/1755-8417-1-4
 80. Gu S, Yuan B, Campbell IM, et al. Alu-mediated diverse and complex pathogenic copy-number variants within human chromosome 17 at p13.3. *Hum Mol Genet.* 2015;24(14):4061-4077. doi:10.1093/hmg/ddv146
 81. Lieber MR. The Mechanism of Human Nonhomologous DNA End Joining. *J Biol Chem.* 2008;283(1):1-5. doi:10.1074/jbc.R700039200
 82. Muentert MD, Forno LS, Hornykiewicz O, et al. Hereditary form of parkinsonism--dementia. *Ann Neurol.* 1998;43(6):768-781. doi:10.1002/ana.410430612
 83. Lee JA, Carvalho CMB, Lupski JR. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell.* 2007;131(7):1235-1247. doi:10.1016/j.cell.2007.11.037
 84. Beck CR, Carvalho CMB, Banser L, et al. Complex genomic rearrangements at the PLP1 locus include triplication and quadruplication. Spinner NB, ed. *PLoS Genet.* 2015;11(3):e1005050. doi:10.1371/journal.pgen.1005050
 85. Liu P, Gelowani V, Zhang F, et al. Mechanism, prevalence, and more severe neuropathy phenotype of the Charcot-Marie-Tooth type 1A triplication. *Am J Hum Genet.* 2014;94(3):462-469. doi:10.1016/j.ajhg.2014.01.017

86. Beutler E, Gelbart T, West C. Identification of Six New Gaucher Disease Mutations. *Genomics*. 1993;15(1):203-205. doi:10.1006/geno.1993.1035
87. Stone DL, Carey WF, Christodoulou J, et al. Type 2 Gaucher disease: the collodion baby phenotype revisited. *Arch Dis Child Fetal Neonatal Ed*. 2000;82(2):F163-6.
http://www.ncbi.nlm.nih.gov/pubmed/10685993. Accessed March 20, 2019.
88. Cozar M, Bembi B, Dominissini S, et al. Molecular characterization of a new deletion of the GBA1 gene due to an inter Alu recombination event. *Mol Genet Metab*. 2011;102(2):226-228. doi:10.1016/j.ymgme.2010.10.004
89. Sidransky E, Lopez G. The link between the GBA gene and parkinsonism. *Lancet Neurol*. 2012;11(11):986-998. doi:10.1016/S1474-4422(12)70190-4
90. Ysselstein D, Shulman JM, Krainc D. Emerging links between pediatric lysosomal storage diseases and adult parkinsonism. *Mov Disord*. February 2019;doi: 10.1002/mds.27631. doi:10.1002/mds.27631
91. Yahalom G, Greenbaum L, Israeli-Korn S, et al. Carriers of both GBA and LRRK2 mutations, compared to carriers of either, in Parkinson's disease: Risk estimates and genotype-phenotype correlations. *Parkinsonism Relat Disord*. December 2018;doi: 10.1016/j.parkreldis.2018.12.014. doi:10.1016/j.parkreldis.2018.12.014
92. Gonzaga-Jauregui C, Harel T, Gambin T, et al. Exome Sequence Analysis Suggests that Genetic Burden Contributes to Phenotypic Variability and Complex Neuropathy. *Cell Rep*. 2015;12(7):1169-1183. doi:10.1016/j.celrep.2015.07.023
93. Nalls MA, Saad M, Noyce AJ, et al. Genetic comorbidities in Parkinson's disease. *Hum Mol Genet*. 2014;23(3):831-841. doi:10.1093/hmg/ddt465
94. Poorkaj P, Nutt JG, James D, et al. parkin mutation analysis in clinic patients with early-

onset Parkinson's disease. *Am J Med Genet.* 2004;129A(1):44-50.

doi:10.1002/ajmg.a.30157

95. Kay DM, Moran D, Moses L, et al. Heterozygous *parkin* point mutations are as common in control subjects as in Parkinson's patients. *Ann Neurol.* 2007;61(1):47-54.

doi:10.1002/ana.21039

Figure Legends

Figure 1: aCGH plots and breakpoint junction sequences of two CNVs involving *SNCA*.

(A) In Subject 3, a 248-kb duplication was identified. In this case, the whole *SNCA* gene was duplicated. The junction sequence (Bottom) is aligned with upstream and downstream reference sequences, with the blue and pink colors indicating their different origins, and the red indicating inserted nucleotides and microhomology. (B) A 1.7-Mb DUP-TRP/INV-DUP rearrangement was identified in the index patient with a known *SNCA* multiplication^{39,40,82}. The x-axis indicates the chromosomal regions surrounding *SNCA*. The y-axis indicates the subject versus control log₂ ratio of the aCGH results, with duplications at 0.58, triplications at 1, and heterozygous deletions at -1 based on theoretical calculations. Red dots in the graph represent probes with log₂ ratio >0.25, black dots with log₂ ratio from 0.25 to -0.25, and green dots with log₂ ratio <-0.25. The normal-duplication-triplication transition regions are magnified in boxes above the plot. The entire *SNCA* gene is triplicated. In addition, a SNP (rs12651181, underlined) was detected close to JCT2.

Figure 2: aCGH plot and breakpoint junction sequence of *GBA* deletion. (A) aCGH plot and (B) junction sequence of a 4.7-kb deletion identified involving *GBA* in patient 1. The deletion (shadowed) encompasses 7 exons of *GBA* (from exon 2 to exon 8). (C) By agarose gel electrophoresis, the amplification of the deleted region (Del) in Subject 1 showed a ~5kb discrepancy compared to a control subject (Ctl), consistent with aCGH findings. PCR showed preferential amplification of the shorter fragment in the Del lane.

Figure 3: aCGH plots and breakpoint junction sequences of *PARK2* CNVs. aCGH plots (left panel) and breakpoint junction sequences (right panel) of CNVs identified involving *PARK2* gene in the cohort. At the top, a schematic gene structure demonstrates the 12 exons of *PARK2*. (A) In subject 20, a 222-kb deletion covering exons 8 and 9 was accompanied by a known pathogenic nonsense mutation c.1358G>A:p.W453* (gnomAD frequency = 0) in exon 12. (B) In subject 11, in addition to a missense variant c.1310C>T:p.P437L (exon 12)) a duplication-normal-duplication (DUP-NML-DUP) was identified. (C) Siblings 21 and 22 share a pathogenic missense variant c.719C>T:p.T240M (in exon 6) and a 178 kb deletion (disrupting exons 5 and 6). (D) In subject 6, a 364-kb duplication encompassed exons 4 to 6. A known pathogenic frameshift mutation c.155delA:p.N52fs was identified in exon 2 (gnomAD frequency = 2.5×10^{-4}). Breakpoint sequencing was not successful in this sample.

Figure 4: *PARK2* and *GBA* ddPCR results of representative subjects. (A) Positive droplet concentrations in 8 subjects are shown. Primer pairs for the 12 exons of *PARK2* and two control genes, *RPPH1* and *TERT*, were used to obtain positive droplet concentrations from PCR in each individual (Supplementary Methods and Supplementary Figure 2). 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 Supplementary Figure 2. Copy

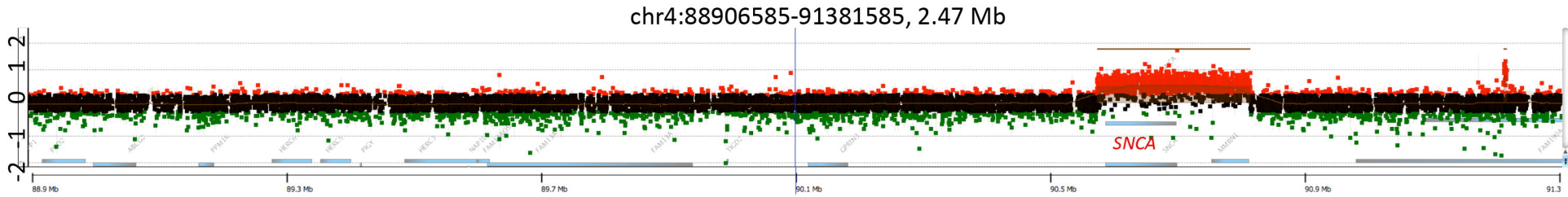
number variants are denoted with asterisks (*). (B) *GBA* and its nearby pseudogene, *GBAP1*, share a high degree of sequence homology, with ddPCR primer pairs for 6 of the 12 exons of *GBA* producing amplicons concurrently from *GBA* and *GBAP1*: *GBA* exons 3, 5, 6, 8, 11, and 12 are color coded to 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 was used to detect potential exonic CNVs in *GBA*. Here, we demonstrate a deletion identified in subject 1, compared to HapMap subject NA10581 and other two 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.

Table 1: SNVs associated with increased PD risk detected via exome sequencing. All indicated SNVs were heterozygous, except *PARK2* c.719C>T:p.T240M, 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). Non-diagnostic (N), heterozygous SNVs were also discovered in *PARK2* (p.R275W) and *DJ-1* (p.A104T). In 2 subjects, SNVs in both *LRRK2* and *GBA* were identified (double heterozygotes): ^A*LRRK2* p.G2019S and *GBA* p.E365K; ^B*LRRK2* p.G2019S and *GBA* p.L483P. ^CInterpretations of c.1310C>T:p.P437L are conflicting ^{94,95}.

Gene	Variant	Diagnostic	Subjects (n)
<u>Dominant</u>			
<i>LRRK2</i>	c.6055G>A:p.G2019S	Y	3 ^{A,B}
<i>LRRK2</i>	c.4321C>G:p.R1441G	Y	1
<i>GBA</i>	c.1093G>A:p.E365K	Y	5 ^A
<i>GBA</i>	c.1223C>T:p.T408M	Y	5
<i>GBA</i>	c.1448T>C:p.L483P	Y	1 ^B
<i>GBA</i>	c.1226A>G:p.N409S	Y	2
<u>Recessive</u>			
<i>PARK2</i>	c.155delA:p.N52fs	Y*	1
<i>PARK2</i>	c.1310C>T:p.P437L ^C	Y*	1
<i>PARK2</i>	c.1358G>A:p.W453*	Y*	1
<i>PARK2</i>	c.719C>T:p.T240M	Y*	2
<i>PARK2</i>	c.823C>T:p.R275W	N	1
<i>DJ-1</i>	c.310G>A:p.A104T	N	1

Figure 1

A



B

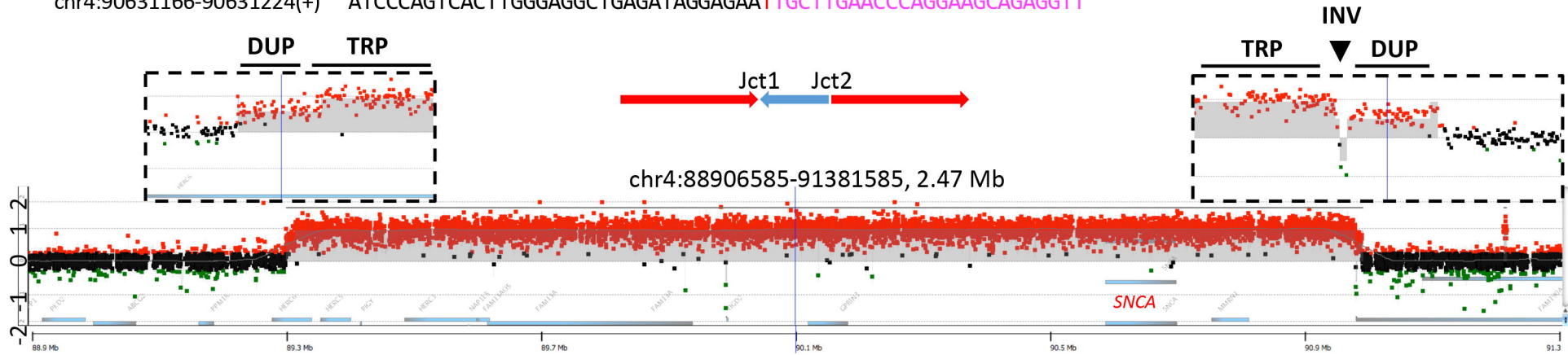
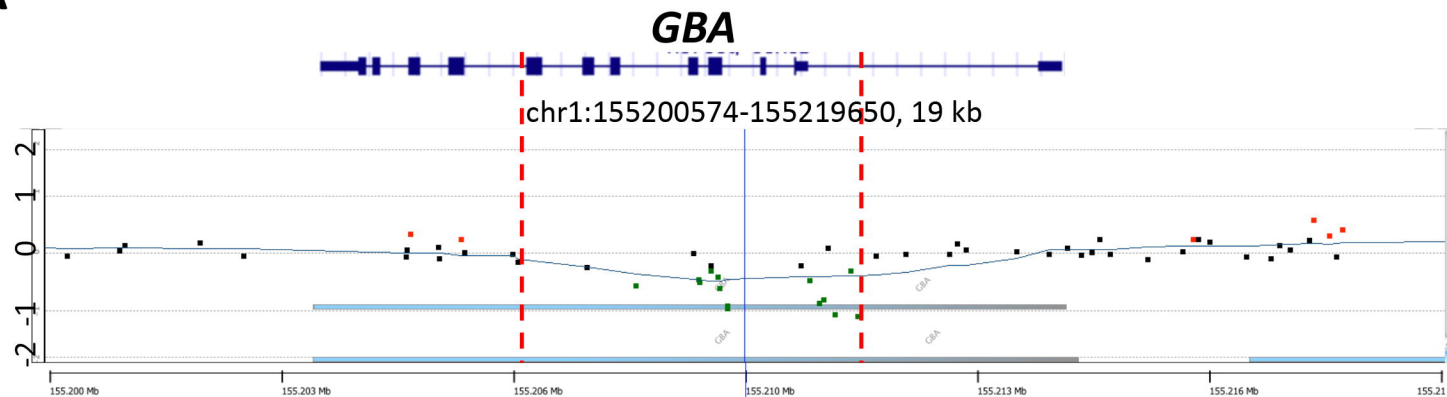


Figure 2**A****B**

chr1:155207041-155207100 TTGGGTAGAGAAATCGCTCTAAGTTTGGGAGCCAAGTCATTTGGATGCTGGATTTGAAGGT
|||||
TTGGGTAGAGAAATCGCTCTAAGTTTGGGAGCCATATCCTCAGTGGACACAAGGAGGAAG

chr1:155211733-155211792 AGTCGTTCAAAAACCAAAGTGTGTACAAAGCCATATCCTCAGTGGACACAAGGAGGAAG
|||||

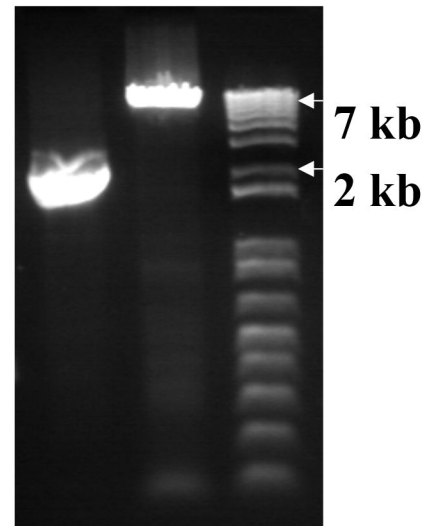
C**Del Ctrl**

Figure 3

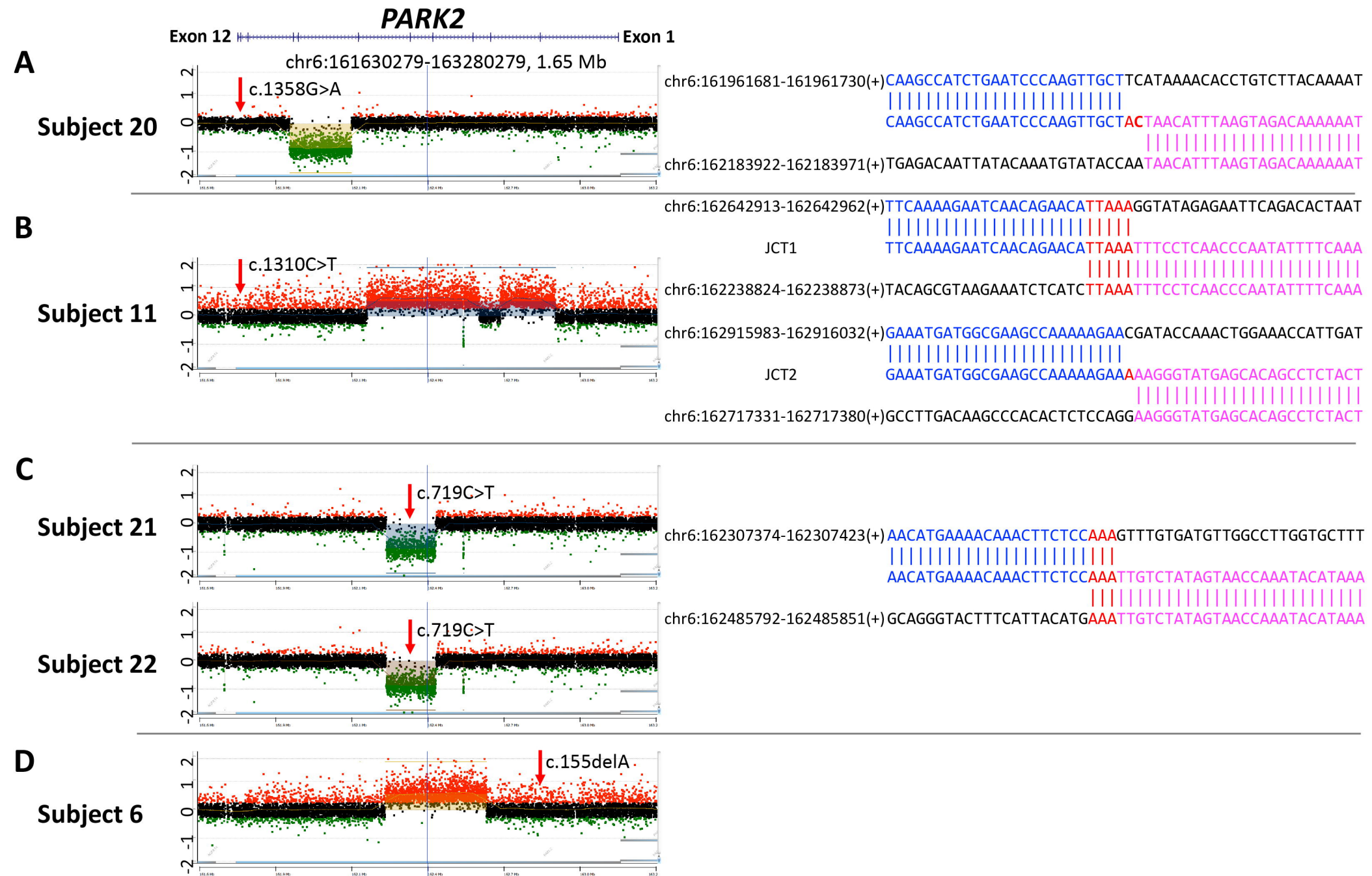
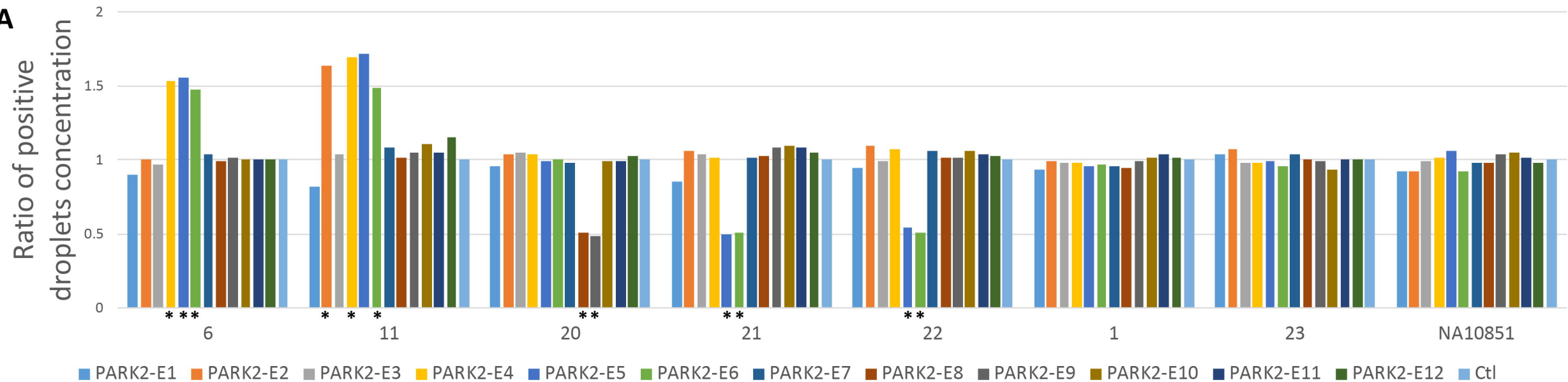
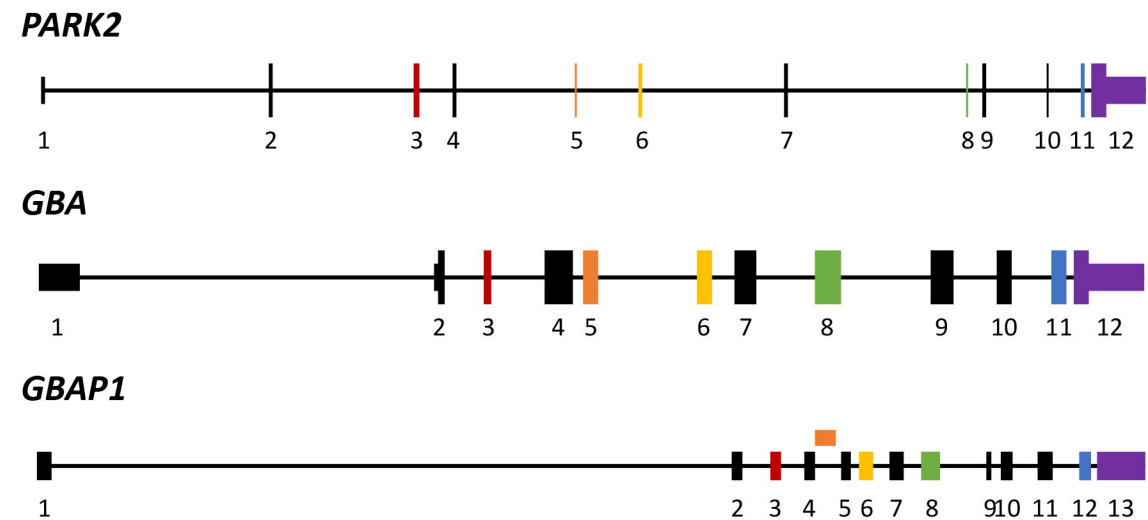


Figure 4**A****B****C**