

relatively short term, the stimulation led to reduction in her torticollis and a better ability to walk. ■

References

1. Kojovic M, Pareés I, Lampreia T, et al. The syndrome of deafness-dystonia: clinical and genetic heterogeneity. *Mov Disord* 2013;28:795-803.
2. van Egmond ME, Kuiper A, Eggink H, et al. Dystonia in children and adolescents: a systematic review and a new diagnostic algorithm. *J Neurol Neurosurg Psychiatry* 2015;86:774-781.
3. Neveling K, Feenstra I, Gilissen C, et al. A post-hoc comparison of the utility of sanger sequencing and exome sequencing for the diagnosis of heterogeneous diseases. *Hum Mutat* 2013;34:1721-1726.
4. Baraitser M, Winter RM. Iris coloboma, ptosis, hypertelorism, and mental retardation: a new syndrome. *J Med Genet* 1988;25:41-43.
5. Verloes A, Di Donato N, Masliah-Planchon J. Baraitser-Winter cerebrofrontofacial syndrome: delineation of the spectrum in 42 cases. *Eur J Hum Genet* 2015;23:292-301.
6. Procaccio V, Salazar G, Ono S, et al. A mutation of beta-actin that alters depolymerisation dynamics is associated with autosomal dominant developmental malformations, deafness, and dystonia. *Am J Hum Genet* 2006;78:947-960.
7. Vidailhet M, Jutras MF, Grabli D, Roze E. Deep brain stimulation for dystonia. *J Neurol Neurosurg Psychiatry* 2013;84:1029-1042.
8. Havránková P, Jech R, Roth J, Urgosik D, Ruzicka E. Beneficial effect of deep brain stimulation of GPi in a patient with dystonia-deafness phenotype. *Mov Disord* 2009;24:465-466.
9. Cif L, Gonzalez V, Garcia-Pracek S, et al. Progressive dystonia in Mohr-Tranebjaerg Syndrome with cochlear implant and deep brain stimulation. *Mov Disord* 2013;28:737-738.

Supporting Data

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

Detection of Genomic Rearrangements From Targeted Resequencing Data in Parkinson's Disease Patients

Nino Spataro, MSc,¹ Ana Roca-Umbert, MSc,¹ Laura Cervera-Carles, MSc,^{2,3} Mònica Vallès, LT,¹ Roger Anglada, LT,⁴ Javier Pagonabarraga, MD, PhD,^{2,3} Berta Pascual-Sedano, MD, PhD,^{2,3,5} Antònia Campolongo, BSc,^{2,3,5} Jaime Kulisevsky, MD, PhD,^{2,3,5} Ferran Casals, PhD,⁴ Jordi Clarimón, PhD,^{2,3*} and Elena Bosch, PhD^{1*}

¹Institute of Evolutionary Biology (CSIC-UPF), Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, Spain ²Department of Neurology, Institut d'Investigacions Biomèdiques Sant Pau-Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, Spain ³Center for Networking Biomedical Research in Neurodegenerative Diseases (CIBERNED), Madrid, Spain ⁴Genomics Core Facility, Universitat Pompeu Fabra, Barcelona Biomedical Research Park (PRBB), Barcelona, Spain ⁵Health Sciences Department, Universitat Oberta de Catalunya, Catalonia, Spain

ABSTRACT

Background: The analysis of coverage depth in next-generation sequencing data allows the detection of gene dose alterations. We explore the frequency of such structural events in a Spanish cohort of sporadic PD cases.

Methods: Gene dose alterations were detected with the eXome-Hidden Markov Model (XHMM) software from depth of coverage in resequencing data available for 38 Mendelian and other risk PD loci in 394 individuals (249 cases and 145 controls) and subsequently validated by quantitative PCR.

Results: We identified 10 PD patients with exon dosage alterations in *PARK2*, *GBA-GBAP1*, and *DJ1*. Additional functional variants, including 2 novel nonsense mutations (p.Arg1552Ter in *LRRK2* and p.Trp90Ter in *PINK1*), were confirmed by Sanger sequencing. This combined approach disclosed the genetic cause of 12 PD cases.

Conclusions: Gene dose alterations related to PD can be correctly identified from targeting resequencing data. This approach substantially improves the detection rate of cases with causal genetic alterations. © 2016 The Authors. *Movement Disorders* published by Wiley Periodicals, Inc. on behalf of International Parkinson and Movement Disorder Society.

Key Words: Parkinson's disease; next generation sequencing; structural variants; XHMM software

Next-generation sequencing (NGS) technologies such as whole-genome, whole-exome, and custom targeting sequencing strategies are increasingly being used to understand the genetic factors underlying both common and rare neurological disorders.^{1,2} Notably, the recent advent of the NGS technologies has been accompanied by the implementation of new bioinformatic tools to detect copy number variants (CNVs) from the depth of read mapping, even after applying

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

*Corresponding authors: Dr. Elena Bosch, Institute of Evolutionary Biology (CSIC-UPF), Department of Experimental and Health Sciences, Universitat Pompeu Fabra, C/Doctor Aiguader 88, 08003 Barcelona, Spain; elena.bosch@upf.edu. Dr. Jordi Clarimón, Department of Neurology, Institut d'Investigacions Biomèdiques Sant Pau-Hospital de la Santa Creu i Sant Pau, C/Sant Antoni M. Claret 167, 08025 Barcelona, Spain; JClarimon@santpau.cat

Funding agencies: This work was supported by the Ministerio de Ciencia e Innovación, Spain (SAF2011-29239), Ministerio de Economía y Competitividad and Fondo Europeo de Desarrollo Regional (FEDER) (SAF2012-35025 and SAF2015-68472-C2-2-R), and by the Direcció General de Recerca, Generalitat de Catalunya (2014SGR-866 and 2014SGR-0235).

Relevant conflicts of interests/financial disclosures: Nothing to report.

Received: 29 June 2016; **Revised:** 12 September 2016; **Accepted:** 27 September 2016

Published online 7 November 2016 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/mds.26845

specific targeted enrichment.³⁻⁷ Most of these new tools are based on the assumption that differences in the depth of coverage among specific genomic regions and across multiple samples can be used as an indicator of the relative number of copies, resulting in a semiquantitative estimation of CNVs. However, as far as we know, this approach has never been attempted in PD.

The main aim of this study was to explore whether CNVs help to explain some sporadic PD patients in which no causative point mutations are found. In particular, we used our own resequencing data of 38 PD-associated genes in 249 PD cases and 145 unrelated controls of European ancestry⁸ to study the presence of structural variants predicted by eXome-Hidden Markov Model (XHMM) software (<https://atgu.mgh.harvard.edu/xhmm/>), which was specifically developed to recover information on copy number variation from normalized read depth data obtained from targeted sequencing experiments.³ Subsequently, all the predicted CNVs were validated through quantitative polymerase chain reaction (PCR). Moreover, detailed frequencies for all potentially functional exon dose alterations detected here and for previously described pathogenic single nucleotide polymorphisms (SNPs) and indels in the same resequencing dataset are provided to understand their relative relevance in PD.

Methods

Participants and Targeted Resequencing Data

Targeted resequencing data from 394 individuals including 249 idiopathic PD cases and 145 unrelated controls of European origin were compiled from Spataro et al. (2015).⁸ Details on eligibility criteria, clinical and demographic features of PD patients and targeting resequencing design are available in Appendix 1 (Supplementary Data).

Coverage and Detection of CNVs

The mean coverage per target and sample was 49.39X, and 91% of the target bases were covered at $\geq 15X$ depth.⁸ The detection of CNVs was performed with the XHMM software, which uses principal component analysis normalization and a hidden Markov model to detect and genotype CNVs from normalized read-depth data from targeted sequencing experiments.³ Phred-scaled quality scores for the CNV events in the inferred intervals ranged from 30 to 99, with a mean of 80 (see Table S1). Gene dose alterations in Mendelian PD genes were then subsequently validated by quantitative PCR (see details in Appendix 2 in the Supplementary Data).

PCR and Sanger Sequencing Validation Analysis

Frameshift deletions and stop gain mutations within the *PARK2*, *LRRK2* and *PINK1* genes plus genomic rearrangements around the *GBA-GBAP1* region were confirmed by PCR and Sanger sequencing analyses (see details in Appendix 3 in the Supplementary Data).

Collapsing Tests

The potential enrichment of exon dosage alterations was tested in the whole set of cases and controls using the full list of CNVs reported in Table S1. CNV enrichment was performed with the VariantTools package (<http://varianttools.sourceforge.net/>), which includes up to 12 different collapsing tests.⁹

Results

Our study of structural genetic alterations across the 38 PD-associated genes previously sequenced through NGS⁸ disclosed a total of 11 structural variants in the *PARK2*, *GBA*, and *DJ1* genes, affecting 10 of 249 PD cases. All of these genomic alterations were predicted by the XHMM software³ (see details in Table S1) and were subsequently validated by quantitative PCR (Figure 1 and Figure S1). CNVs were also detected in 2 of 145 controls in a heterozygous state for the recessive gene *PARK2* and the *GBA-GBAP1* region. The clinical features of the PD samples in which structural variants were found are provided in Appendix 1 (Supplementary Data). Additional functional sequence alterations detected in the same dataset are listed in Table 1 (see Table S2 for complete genotypes).

Structural Variants in the *PARK2* Gene

We identified 4 different exon deletions and 1 exon duplication in the *PARK2* gene, affecting a total of 6 PD cases (Figure 1A). All of them have already been described in PD cases.^{10,11} The 2 largest *PARK2* deletions spanned from exon 3 to 6 and from exon 2 to 4 and were found in heterozygosis in Cas241 and Cas20, respectively. These 2 PD cases are examples of compound heterozygotes because both individuals also present different frameshift mutations in exon 2, causing a premature stop codon: Cas241 is heterozygote for a dinucleotide deletion (p.Gln34Argfs), whereas Cas20 is hemizygote for a single-nucleotide deletion (p.Asn52Metfs; see Table 1, Table S2, and Figure S2). The *PARK2* deletion comprising exons 3 to 4 was found in 2 different PD cases: Cas57, which is homozygote for the deletion, and Cas246, which is an additional *PARK2* compound heterozygote also carrying the p.Asn52Metfs mutation. Finally, Cas232 is found homozygote for a *PARK2* deletion affecting exon 2, and Cas148 is homozygote for a duplication affecting exon 3. As expected, early age at onset (before 45

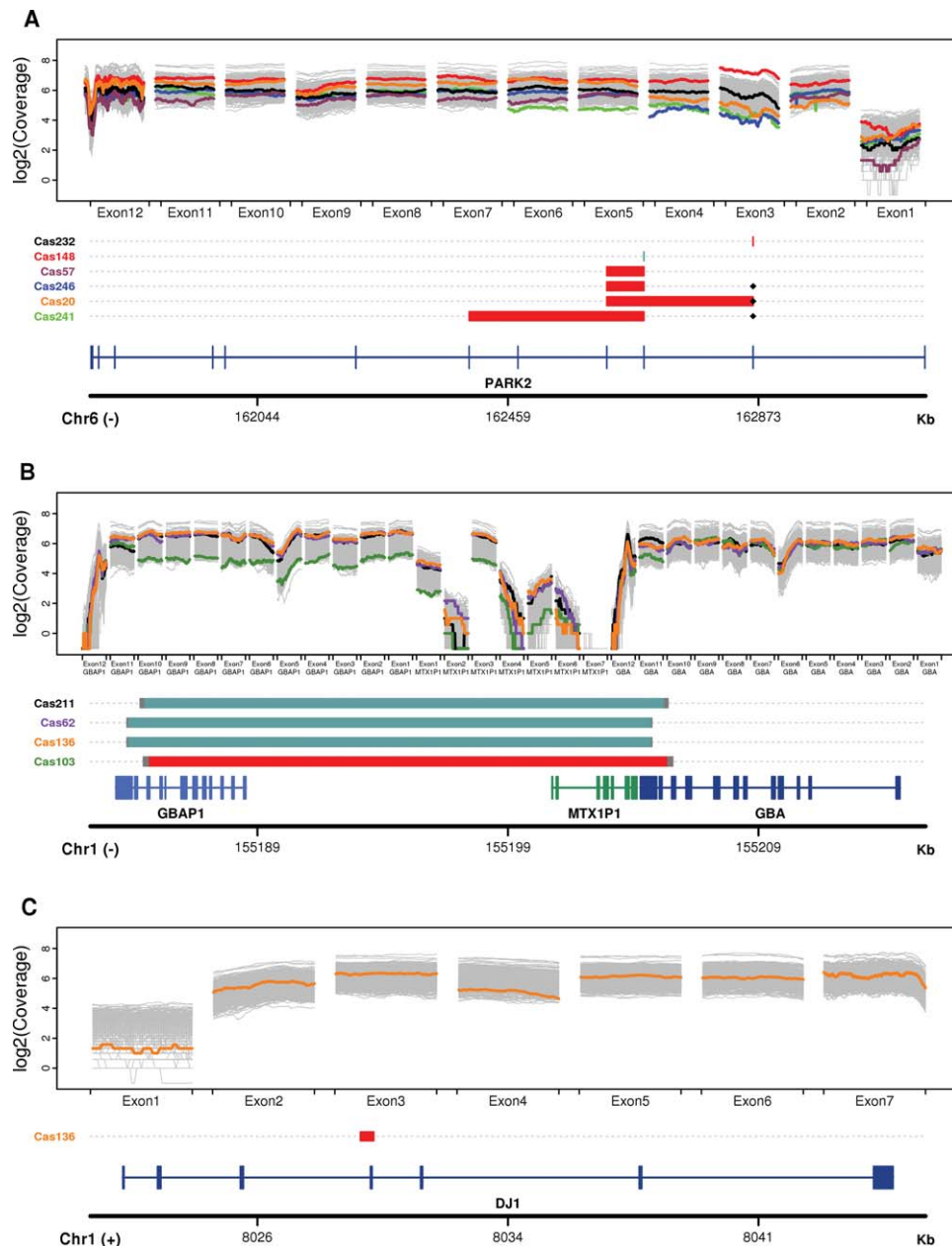


FIG. 1. Detection of copy number variation: **(A)** *PARK2* gene, **(B)** *GBA-GBAP1* region, and **(C)** *DJ1* gene. In the upper panel, sequencing depth of coverage for those samples inferred to carry copy number variants by the eXome-Hidden Markov Model (XHMM) software (<https://atgu.mgh.harvard.edu/xhmm/>) is indicated in colors along each exon; the color matches that of the individual label on the Y axis of the bottom panel. In all other samples, the gray background represents sequencing depth of coverage. In the bottom panel, a schematic representation of the corresponding validated copy number variants is presented for each gene. Red bars, deletion; light blue bars, duplication; black diamonds, frameshift indels. The bottom track represents a schematic representation of the gene structure.

years of age) was found in 4 of the 6 patients with structural variants in *PARK2*.

Structural Variants in the *GBA-PGBA1* Region

The high homology between *GBA* and its neighboring pseudogene (*GBAP1*), which share 96% of sequence identity, not only explains several gene-pseudogene rearrangements and gene-conversion events¹² but also complicates the analysis of the whole region.¹³ Our analysis disclosed 4 individuals

presenting CNVs along the *GBA-GBAP1* region (see Figure 1B). Cas103, which has already been described in Setó-Salvia et al. (2012),¹⁴ is heterozygote for a recombinant deletion known as the Rec-Ncil allele (where *GBAP1* exons 1 to 10 and *GBA* exons 11 and 12 are deleted). In contrast, Cas211, Cas62, and Cas136 are heterozygotes for 2 different duplications along the same region. Because exons 11 and 12 in the gene and the pseudogene are nearly identical, we could not verify by quantitative PCR the exact limits of the *GBA-GBAP1* rearrangements detected (Figure

TABLE 1. Summary of pathogenic alterations found in Mendelian genes related to PD.

Gene	Mutation type	DNAchange	Proteinchange	Cases (2N = 498)	Controls (2N = 290)	Sample ID (coverage)
<i>PARK2</i> (M Recessive)	Missense	c.802G>A ¹	p.Arg234Gln	3	0	Cas74 (58), Cas172(126), Cas214 (75)
	Frameshift	c.154delA ²	p.Asn52Metfs	3	1	Cas211 (31), Cas246 ^{ab} (32), Cas20 ^{ab} (30), Con142(45)
	Exon deletion	Ex3-4del		3	0	Cas57 ^c (0) ^d , Cas246 (20) ^d
	Exon deletion	Ex2del		2	0	Cas232 ^c (0) ^d
	Missense	c.1244C>A ³	p.Thr415Asn	1	0	Cas211 (76)
	Missense	c.574A>C ⁴	p.Met192Leu	1	0	Cas76 (72)
	Exon deletion	Ex2-4del		1	0	Cas20 (34) ^d
	Exon deletion	Ex3-6del		1	0	Cas241 (24) ^d
	Exon duplication	Ex3dup		2	0	Cas148 ^c (149) ^d
	Frameshift	c.220-221insGT ⁵	p.Trp74Cysfs	1	0	Cas11 (29)
	Frameshift	c.101-102delAG ⁶	p.Gln34Argfs	1	0	Cas241 ^{ab} (25)
<i>LRRK2</i> (M Dominant)	Missense	c.6055G>A ⁷	p.Gly2019Ser	3	0	Cas213 (126), Cas226 (39), Cas113 (108)
	Stop gain	c.4654C>T ⁸	p.Arg1552Ter	1	0	Cas55 ^a (81)
<i>DJ1</i> (M Recessive)	Exon deletion	Ex4del		1	0	Cas136 (36) ^d
<i>PINK1</i> (M Recessive)	Stop gain	c.270G>A ⁸	p.Trp90Ter	2	0	Cas154 ^{ac} (2)
	Stop gain	c.1366C>T ⁹	p.Gln456Ter	1	0	Cas194 (30)
<i>GBA</i>		RecNcil		1	0	Cas103 (61) ^d

¹rs144032774 (C/T), ²rs754809877 (T/-), ³rs778125254 (G/T), ⁴rs9456735 (T/G), ⁵rs746646126 (-/CA), ⁶rs55777503 (CT/-), ⁷rs34637584 (G/A), ⁸Not reported before, ⁹rs45539432 (C/T), ^aConfirmed with Sanger sequencing, ^bHemizygous, ^cHomozygous, ^dAverage coverage across duplication/deletion. Abbreviations: 2N, number of chromosomes.

1B). Subsequent PCR and sequencing analysis allowed us to confirm that Cas211 is heterozygote for the reciprocal product of the Rec-Ncil deleted allele (exon 11 and 12 of the *GBA* gene and *GBAP1* exons 1 to 10 are duplicated), whereas Cas62 and Cas136 are heterozygotes for a duplication affecting the 3'untranslated region of the *GBA* gene and most of the *GBAP1* pseudogene (see details in Figure S3 and Table S3). In addition to the *GBA-GBAP1* duplication rearrangement, Cas62 also carries the p.Asp370Ser mutation described to increase risk for late-onset PD¹⁵ and known to be the most common causal mutation for Gaucher's disease in Ashkenazi Jews¹⁶ in a heterozygous state (Table S2).

Structural Variants in the *DJ1* Gene

Among all analyzed participants, only 1 PD patient (Cas136) carried a heterozygous deletion comprising the whole exon 4 of the *DJ1* gene (Figure 1C). As far as we are aware, this alteration has never been reported before.^{10,11} Interestingly, the same individual is also heterozygote for 1 duplication in the *GBA-GBAP1* region. Therefore, Cas136 could be a particular case of a double heterozygote for genomic rearrangements occurring in 2 different PD loci.

Other Functional Mutations and Enrichment of Rare Variation

Besides the detected structural variants, other functional alterations (including several known PD

Mendelian mutations) had been previously detected in the same dataset.⁸ Frequencies and details are presented in Table 1 and Appendix 4 (Supplemental Data). Here, we further validated particular genotypes by Sanger sequencing in some PD cases to confirm genotypes in regions with low coverage depth as well as novel potential causative variants. Notably, we confirmed the identification of 2 stop-gain mutations, to our knowledge previously unreported, having checked the PDmutDB,^{10,11} 1000 Genomes Project,¹⁷ ExAC database (<http://exac.broadinstitute.org/>, accessed April 26, 2016) and the dbSNP database¹⁸: p.Arg1552Ter in *LRRK2*, which was found in 1 heterozygote carrier (Cas55), and p.Trp90Ter in *PINK1*, which was found in homozygosis (Cas154) and could also represent a new causal variant for PD (see Figure S2).

In the same dataset, we have previously shown that PD cases displayed significantly higher proportions of rare (minor allele frequency (MAF) <1%) code-altering (nonsynonymous SNPs, nonsense mutations, and coding indels) variants than controls on Mendelian genes.⁸ Notably, when performing the same type of collapsing analyses, considering only the exon dosage alterations identified here, PD cases also show significant enrichment of CNV in PD Mendelian genes when compared with controls (*P* value <.05 in 10 of 12 tests; Figure S4).

Discussion

Our analysis demonstrates the usefulness of NGS for discovering different types of variants with a potential

role in human disease. Among the detected inactivating variants, we not only report point mutations and small indels but also different exon dosage variants already known to be involved in PD aetiology. Notably, all predicted CNVs were subsequently confirmed by quantitative PCR, suggesting that the analysis of coverage from resequencing data across multiple individuals provides valuable information for the identification of exon dosage variants. It should be noted, however, that the sensitivity of the XHMM software could not be evaluated with this research design and that our study was focused on a limited set of candidate genes. Thus, an unknown number of CNVs could remain undiscovered in this set of patients.

Whereas the analysis of functional SNP variation and indels in Mendelian genes related to PD allowed us to identify putative causative variants for 6 PD cases (Cas211, Cas213, Cas226, Cas113, Cas55, and Cas154), the joint analysis of these inactivating variants, together with the exon dose alterations detected here, probably helps to explain the disease phenotype of 6 additional PD cases (Cas246, Cas20, Cas57, Cas232, Cas148, Cas241) in our Spanish cohort of 249 PD cases (2.4%). Thus, as demonstrated in this dataset, CNVs in the form of exon dose alterations are at least as important as indels and other functional SNP variations in PD Mendelian genes when explaining the phenotype of apparently sporadic PD cases. Given the recognized role of structural variants in several neurodegenerative disorders and because many NGS-based projects with large numbers of individuals are currently underway to study rare and common variant association, efforts should be made to integrate the analysis of potential CNVs in these new datasets. ■

Acknowledgments: The authors are grateful to all participants in the study. This work was supported by the Ministerio de Ciencia e Innovación, Spain (SAF2011-29239), Ministerio de Economía y Competitividad, Spain, and Fondo Europeo de Desarrollo Regional (FEDER) (SAF2012-35025 and SAF2015-68472-C2-2-R), and by the Direcció General de Recerca, Generalitat de Catalunya (2014SGR-866 and 2014SGR-0235). E.B. is the recipient of an Institució Catalana de Recerca i Estudis Avançats (ICREA) Academia Award.

References

1. Keogh MJ, Chinnery PF. Next generation sequencing for neurological diseases: new hope or new hype? *Clin Neurol Neurosurg* 2013;115(7):948-953. <http://www.sciencedirect.com/science/article/pii/S0303846712005227>
2. Handel AE, Disanto G, Ramagopalan SV. Next-generation sequencing in understanding complex neurological disease. *Expert Rev Neurother* 2013;13(2):215-227. <http://dx.doi.org/10.1586/ern.12.165>
3. Fromer M, Moran JL, Chambert K, et al. Discovery and statistical genotyping of copy-number variation from whole-exome

- sequencing depth. *Am J Hum Genet* 2012;91(4):597-607. <http://dx.doi.org/10.1016/j.ajhg.2012.08.005>
4. Love MI, Mysičková A, Sun R, Kalscheuer V, Vingron M, Haas SA. Modeling read counts for CNV detection in exome sequencing data. *Stat Appl Genet Mol Biol* 2011;10(1):52.
5. Krumm N, Sudmant P, Ko A. Copy number variation detection and genotyping from exome sequence data. *Genome Res* 2012;22(8):1525-32. <http://genome.cshlp.org/content/22/8/1525.short>
6. Plagnol V, Curtis J, Epstein M, et al. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics*. 2012;28(21):2747-2754.
7. Samarakoon PS, Sorte HS, Stray-Pedersen A, Rødningen OK, Rognes T, Lyle R. cnvScan: a CNV screening and annotation tool to improve the clinical utility of computational CNV prediction from exome sequencing data. *BMC Genomics* 2016;17:51. <http://dx.doi.org/10.1186/s12864-016-2374-2>
8. Spataro N, Calafell F, Cervera-Carles L, et al. Mendelian genes for Parkinson's disease contribute to the sporadic forms of the disease. *Hum Mol Genet* 2014;24(7):2023-2034. <http://www.hmg.oxfordjournals.org/cgi/doi/10.1093/hmg/ddu616>
9. San Lucas FA, Wang G, Scheet P, Peng B. Integrated annotation and analysis of genetic variants from next-generation sequencing studies with variant tools. *Bioinformatics* 2012;28(3):421-422.
10. Nuytemans K, Theuns J, Cruts M, Van Broeckhoven C. Genetic etiology of Parkinson disease associated with mutations in the SNCA, PARK2, PINK1, PARK7, and LRRK2 genes: a mutation update. *Hum Mutat* 2010;31(7):763-780. <http://dx.doi.org/10.1002/humu.21277>
11. Cruts M, Theuns J, Van Broeckhoven C. Locus-specific mutation databases for neurodegenerative brain diseases. *Hum Mutat* 2012;33(9):1340-1344. <http://dx.doi.org/10.1002/humu.22117>
12. Tayebi N, Stubblefield BK, Park JK, et al. Reciprocal and nonreciprocal recombination at the glucocerebrosidase gene region: implications for complexity in Gaucher disease. *Am J Hum Genet* 2003;72(3):519-534.
13. Horowitz M, Wilder S, Horowitz Z, Reiner O, Gelbart T, Beutler E. The human glucocerebrosidase gene and pseudogene: structure and evolution. *Genomics* 1989;4(1):87-96. <http://www.sciencedirect.com/science/article/pii/0888754389903194>
14. Setó-Salvia N, Pagonabarraga J, Houlden H, et al. Glucocerebrosidase mutations confer a greater risk of dementia during Parkinson's disease course. *Mov Disord* 2012;27(3):393-399. <http://dx.doi.org/10.1002/mds.24045>
15. Hu FY, Xi J, Guo J, et al. Association of the glucocerebrosidase N370S allele with Parkinson's disease in two separate Chinese Han populations of mainland China. *Eur J Neurol* 2010;17(12):1476-1478.
16. Zimran A, Gross E, West C, Sorge J, Kubitz M, Beutler E. Prediction of severity of Gaucher's disease by identification of mutations at DNA level. *Lancet* 1989;334(8659):349-352. <http://www.sciencedirect.com/science/article/pii/S0140673689905369>
17. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012;491(7422):56-65.
18. Sherry ST, Ward M-H, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001;29(1):308-311. <http://nar.oxfordjournals.org/content/29/1/308.abstract>

Supporting Data

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