



# Investigation of Somatic Mutations in Human Brains Targeting Genes Associated With Parkinson's Disease

Melissa Leija-Salazar<sup>1</sup>, Alan Pittman<sup>2</sup>, Katya Mokretar<sup>1</sup>, Huw Morris<sup>1</sup>, Anthony H. Schapira<sup>1</sup> and Christos Proukakis<sup>1\*</sup>

<sup>1</sup> Department of Clinical and Movement Neurosciences, University College London Queen Square Institute of Neurology, London, United Kingdom, <sup>2</sup> Genetics Research Centre, Molecular and Clinical Sciences, St George's University of London, London, United Kingdom

**Background:** Somatic single nucleotide variant (SNV) mutations occur in neurons but their role in synucleinopathies is unknown.

**Aim:** We aimed to identify disease-relevant low-level somatic SNVs in brains from sporadic patients with synucleinopathies and a monozygotic twin carrying *LRRK2* G2019S, whose penetrance could be explained by somatic variation.

**Methods and Results:** We included different brain regions from 26 Parkinson's disease (PD), one incidental Lewy body, three multiple system atrophy cases, and 12 controls. The whole SNCA locus and exons of other genes associated with PD and neurodegeneration were deeply sequenced using molecular barcodes to improve accuracy. We selected 21 variants at 0.33–5% allele frequencies for validation using accurate methods for somatic variant detection.

**Conclusions:** We could not detect disease-relevant somatic SNVs, however we cannot exclude their presence at earlier stages of degeneration. Our results support that coding somatic SNVs in neurodegeneration are rare, but other types of somatic variants may hold pathological consequences in synucleinopathies.

**Keywords:** SNCA, synuclein, Parkinson's disease, somatic mutation, targeted sequencing, synucleinopathies, molecular barcodes

## OPEN ACCESS

### Edited by:

Mathias Tøft,  
University of Oslo, Norway

### Reviewed by:

Kenya Nishioka,  
Juntendo University, Japan  
Georgia Xiomerisiou,  
University of Thessaly, Greece

### \*Correspondence:

Christos Proukakis  
c.proukakis@ucl.ac.uk

### Specialty section:

This article was submitted to  
Neurogenetics,  
a section of the journal  
Frontiers in Neurology

**Received:** 07 June 2020

**Accepted:** 22 September 2020

**Published:** 22 October 2020

### Citation:

Leija-Salazar M, Pittman A, Mokretar K, Morris H, Schapira AH and Proukakis C (2020) Investigation of Somatic Mutations in Human Brains Targeting Genes Associated With Parkinson's Disease. *Front. Neurol.* 11:570424. doi: 10.3389/fneur.2020.570424

## INTRODUCTION

Synucleinopathies are disorders characterized by the pathological aggregation of  $\alpha$ -synuclein (1). Among synucleinopathies, Parkinson's disease (PD) is the commonest disorder and is characterized predominantly by neurodegeneration of dopaminergic neurons in substantia nigra (SN) (2, 3). Somatic variation occurs in human brain and its role in neurodegeneration has started to be explored (4). Current estimations of the occurrence of somatic variants in human brains suggest that single nucleotide variants (SNVs, or "point mutations") could be the most prevalent form (5, 6). Somatic SNVs are reported to increase with age, where large genes or transcriptionally active genomic regions appear to be susceptible (7). Somatic SNVs in coding regions of genes associated with synucleinopathies could contribute directly to these disorders, depending on the amount of affected cells and mechanisms of spread of the aetiological agent [see review (8)]. The study of somatic SNVs has been facilitated by the latest technological improvements. Compared to single-cell studies, bulk-sequencing offers a cost-effective strategy to study somatic variation across

tissues and brain regions of multiple individuals. The error rate of bulk-sequencing at low allele frequencies (AF) can be reduced by using molecular barcodes (9). In this study, we used targeted sequencing in PD-associated genes from post-mortem human brains aimed for the detection of pathogenic somatic SNVs.

## METHODS

Samples were obtained from the Parkinson's UK and Queen Square brain banks. Patients gave informed consent and the study was approved by the local ethics committee. We evaluated 66 samples from multiple brain regions and three matched-blood samples, derived from 42 individuals with the following conditions: 26 PD, 12 control, three Multiple system atrophy (MSA), and one Incidental Lewy Body case (**Supplementary Table 1**). PD cases were sporadic, except for case 18, a manifesting *LRRK2* G2019S carrier, whose identical twin was non-penetrant (10) and somatic variation was suggested as an explanation for the discordance in the development of PD (11). We did not include other monogenic cases, as we did not have access to their brain tissue. The mean and standard deviation for onset age was  $62.0 \pm 11.1$  years, and for disease duration  $10.1 \pm 7.0$  years. This calculation excludes case 18, whereas these were not available.

We used a previously reported protocol for genomic DNA extraction (12) and the Haloplex<sup>HS</sup> method to prepare sequencing libraries. Details about the generation of artificial mosaics, the sequencing panel design, the customization of library preparation and bioinformatic analysis are provided in **Supplementary Table 2** and **Supplementary Figure 1**.

For amplicon sequencing, primers were designed with Primer3Plus (13) to generate amplicons larger than 300 bp, targeting the variants of interest at  $>50$  bp away from the primer annealing sites. Amplicons belonging to the same sample were pooled together before Nextera XT library preparation, following manufacturer instructions. Samples were pooled equimolarly before sequencing using a MiSeq v3 kit (600 cycles). The bioinformatic analysis is described in **Supplementary Figure 2**. Droplet digital PCR (ddPCR) assays were designed using Primer3plus, according to manufacturer. Bulk DNA from putamen, occipital, frontal cortex, and cingulate gyrus was used for this analysis. The ddPCR conditions are described in **Supplementary Table 3**. Data analysis was performed in QuantaSoft Pro v1.0 following Bio-Rad guidelines.

## RESULTS

### Validation of the Methodology

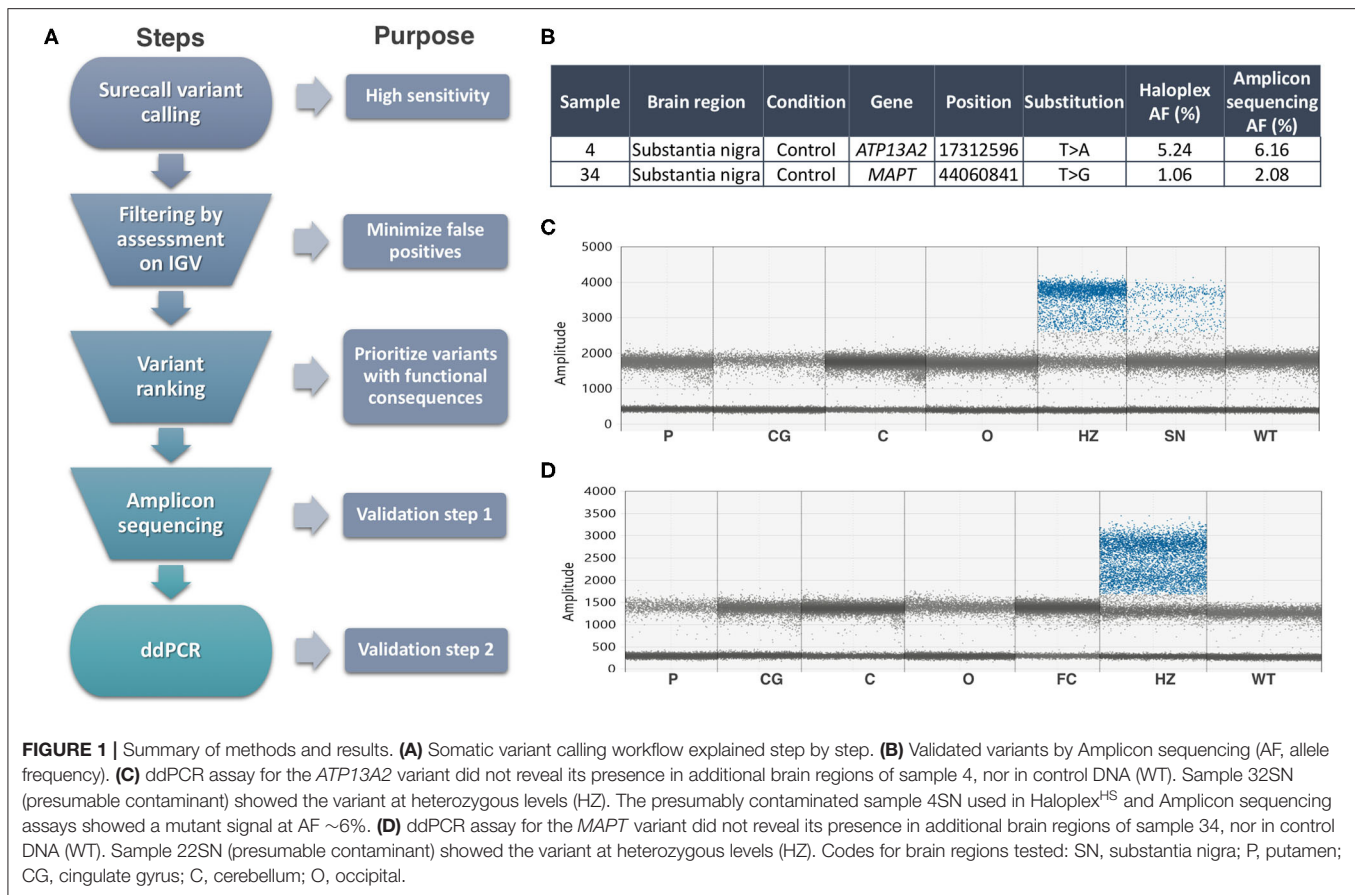
"Artificial mosaics" were used to estimate the variant detection limit, sensitivity and false positive and negative rates. We were expecting 37 variants to be present within regions covered in artificial mosaics. We detected 95% of these variants at 1% AF and 87% at 0.5% AF (supplementary results).

We aimed to reduce to a minimum false positives at lower AF levels. We firstly counted 'Potential false positives' (PFP) in artificial mosaics at different AF thresholds. PFP comprised SNVs not recorded as expected mosaic variants, nor reported in dbSNP

(14). We observed  $1.2\times$  more PFP when the minimum AF was lowered from 1% to 0.5% (**Supplementary Figure 3**). Surecall showed greater sensitivity when compared to other variant callers (**Supplementary Figure 3**). To increase the specificity of our variant calling analysis, we filtered false positives visually, using fixed criteria to discard errors (**Supplementary Figure 4**). Surecall variants in mosaic 0.25% (at AF = 0.25–5%) were analyzed on IGV. From the 114 variants analyzed, visual analysis could not discard 4 false positives. The highest AF was reported as 0.32%, therefore we set our detection limit at 0.33%. This filter allowed us to discard numerous false positives, but also increased the false negative rate. In the artificial mosaic sample carrying variants at 0.5%, Surecall detected 78% of the expected variants. After visual inspections, 46% of the expected variants remained, and false positives were completely discarded. The most common reason to filter real variants was their presence in only one paired-read orientation (strand-bias; **Supplementary Figure 4B**).

### Sample Analysis

We focused on the substantia nigra, and sequenced DNA from 42 samples (including 12 controls). Where available, we also analyzed DNA from other sources from the same individuals (**Supplementary Table 1**): frontal cortex (13, including two controls), cerebellum (11, including 1 control), and blood from three. An explanation of our analysis is summarized in **Figure 1A**. On the Haloplex<sup>HS</sup> step, all samples were sequenced at an average  $2,541\times$ . We focused on the detection of coding SNVs not reported before as common SNPs (population frequencies  $< 1\%$ ), to reduce the risk of calling low-level variants arising due to contamination. Thirty-one variants in 23 samples passed the filtering step, but most of the variants detected (24 out of 31) had an average AF of 0.45%, close to the detection limit of our analysis. Twenty-one variants in 18 samples were prioritized for validation, based on a ranking scale to select variants with a predictable role in disease (**Supplementary Table 4**). We generated amplicons to target the prioritized variants and sequenced those at even higher coverage (mean =  $14,883\times$ ). To account for possible sequencing errors at the genomic positions of interest, we compared the amplicons from the interrogated sample with amplicons from controls (a commercial reference DNA and six samples showing a candidate variant in other parts of the genome). Two variants in samples 4SN and 34SN were validated, as these were detected at AFs close to the original analysis, and significantly different from the sequencing errors in controls (**Figure 1B**). The variants were further confirmed by Mutect2 paired-analysis, using the reference DNA as a normal sample. However, these variants corresponded to rare heterozygous SNPs present in samples from our study. SN tissue was not available for further validation, but the AF at which the variants were detected was an indicative that the variants might be present in other brain regions when real (15). ddPCR did not reveal the variants in the brain regions tested (**Figures 1C,D**). In one of the assays, the presumably contaminated DNA was still available and the variant was confirmed only in this sample (**Figure 1C**). To further examine cross-contamination, we recorded all mosaic variants from Surecall in 4SN, 34SN and control 1 (a sample used for demonstration purposes) at



AF similar to the variants of interest. The mosaic variants were compared to germline variants from samples where the contamination was suspected to come from (in the case of control one, a non-related sample or control two). While control one showed fewer mosaic variants, not matched with control two germline variants, the presumably contaminated samples showed numerous mosaic variants matched with germline variants from samples where the contamination came from ( $p < 0.0001$ , linear regression; **Supplementary Figure 5**).

## DISCUSSION

Previous work from our group could not detect somatic SNVs in *SNCA* exons at AF above 5% in cerebellum, frontal cortex and SN of sporadic PD patients (16). In this study, we expanded our search to other PD-genes. We excluded as many cases as possible with long disease duration and late-onset, as somatic variants playing a role in disease are hypothesized to be less likely to occur in these cases (16, 17). We included a patient carrying a *LRRK2* G2019S mutation, who had a phenotypically discordant monozygotic twin and where somatic variation could have played a role in penetrance (11). We used a highly sensitive approach to detect low-level variants in the genes of interest, by firstly combining deep sequencing coverage and molecular barcodes, followed by amplicon sequencing at higher coverage and ddPCR as validation steps (18). We could not detect somatic

SNVs in PD-associated genes at AF higher than 0.33%. Similar to our results, a recent report could not identify somatic SNVs at AF above 0.5% in familial PD-genes from brains with Lewy body disorders ( $n = 20$ ), using similar methodologies and higher sequencing coverage (19). Previous studies using Haloplex<sup>HS</sup> reported variant detection at AF above 0.2%, further supporting that our analysis was close to the detection limits of this methodology (19–21). We focused on refining the analysis to mainly discard false positives. Our filtering criteria were tailored to discard sequencing artifacts, similar to other studies using Haloplex and common sequencing datasets (22–26). Advantages of visual analysis are the comprehensive analysis for each variant, easy implementation across datasets; however, it can become labor-intensive. Our results demonstrate the difficulties of SNV detection at low AF, due to low-level contamination and false positives, even when using molecular barcodes.

Challenges of somatic variant studies are not only technical, but also related to the stochastic nature of the variants. According to a previous hypothesis where neurons carrying somatic variation may be the most vulnerable and first to degenerate, we selected for patients with disease duration as short as possible (~10 years) (16). When studying neurodegeneration in post-mortem brains, only the latest stages of the disease are being portrayed and, perhaps, events involved in disease development are missed. Conversely, if somatic SNVs arise post-mitotically in an age-dependent manner (4), detailed studies at different age groups are required. Furthermore, as we focused on the SN,

and only had access to DNA from other brain regions or blood in a few cases, we have not provided a detailed assessment of these. The use of patient-derived cell lines or animal models could also be considered. We are not aware of studies of somatic mutations in such samples, but PD patient fibroblasts have inefficient DNA repair, specifically the nucleotide excision repair (NER) pathway, and mice with a mutation compromising NER have dopaminergic pathology (27).

Our data combined with work discussed above, suggest that coding somatic SNVs in PD-associated genes are uncommon. In Alzheimer's disease, two brain somatic SNVs were found in 72 sporadic AD-patients (28). When using molecular barcodes, two brain somatic SNVs were found in AD-associated genes of 98 patients (29), whereas no somatic SNVs in familial AD-genes were found in 20 patients (19). Somatic SNVs in *APP* were reported in AD in the context of the novel mechanism of recombination leading to "genomic cDNA" (30). Recently, 14 out of 52 AD-patients analyzed by deep exome sequencing harbored exonic somatic mutations in genes involved in tau phosphorylation, but not familial AD genes (31). This contrasts with somatic CNVs, with *SNCA* gains in PD nigral dopaminergic and cortical neurons (32, 33).

In summary, our study could not detect coding somatic SNVs at AF above 0.33% when analyzing PD-associated genes from brain samples. Reaching lower AF to detect late somatic variant events using bulk-tissue requires an even larger sequencing effort, and it is complicated by the common presence of contamination and sequencing errors. Sequencing of dopaminergic single-nuclei should give enough resolution to describe somatic variants in cells mainly affected by PD (dopaminergic neurons). Additional studies can be aimed to explore other types of somatic variations or other mechanisms by which somatic SNVs outside PD-associated genes could play detrimental roles in neurodegeneration.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ebi.ac.uk/ena>, PRJEB36518.

## REFERENCES

- Goedert M, Jakes R, Spillantini MG. The synucleinopathies: twenty years on. *J Parkinsons Dis.* (2017) 7:S51–69. doi: 10.3233/JPD-179005
- Rocca WA. Comment the burden of Parkinson's disease: a worldwide perspective. *Lancet Neurol.* (2018) 17:928–9. doi: 10.1016/S1474-4422(18)30355-7
- Mullin S, Schapira AHV. Pathogenic mechanisms of neurodegeneration in Parkinson disease. *Neurol Clin.* (2015) 33:1–17. doi: 10.1016/j.ncl.2014.09.010
- Lodato MA, Walsh CA. Genome aging : somatic mutation in the brain links age-related decline with disease and nominates pathogenic mechanisms. *Hum Mol Genet.* (2019) 28:R197–206. doi: 10.1093/hmg/ddz191
- McConnell MJ, Moran JV, Abyzov A, Akbarian S, Bae T, Cortes-Ciriano I, et al. Intersection of diverse neuronal genomes and neuropsychiatric disease: the brain somatic mosaicism network. *Science.* (2017) 356:eaal1641. doi: 10.1126/science.aal1641
- Bae T, Tomasini L, Mariani J, Zhou B, Roychowdhury T, Franjic D, et al. Different mutational rates and mechanisms in human cells at pregastrulation and neurogenesis. *Science.* (2017) 555:aan8690. doi: 10.1126/science.aan8690
- Lodato MA, Rodin RE, Bohrsen CL, Coulter ME, Barton AR, Kwon M, et al. Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science.* (2018) 359:555–9. doi: 10.1126/science.aao4426
- Leija-Salazar M, Piette CL, Proukakis C. Somatic mutations in neurodegeneration. *Neuropathol Appl Neurobiol.* (2018) 44:267–85. doi: 10.1111/nan.12465

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by NRES Committee central—London. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

ML-S conducted the experiments, analyzed the data, wrote, revised, and submitted this manuscript. AP participated in the experimental design and data analysis. KM participated in the experimental design and performed initial experiments. HM and AS participated in the design of the study. CP conceived and designed the study, contributed to interpret the data and revised the final version of the manuscript. All authors read and approved the final version of the manuscript.

## FUNDING

CP received funding from the Michael J. Fox Foundation for Parkinson's disease research and MS was partly funded by CONACYT. AS was funded by the UK Medical Research Council, and the Kattan Trust. Queen Square Brain Bank is supported by the Reta Lila Weston Institute for Neurological Studies and the Medical Research Council UK. The Parkinson's UK Tissue Bank is funded by Parkinson's UK, a charity registered in England and Wales (258197) and in Scotland (SC037554).

## ACKNOWLEDGMENTS

We are grateful to all individuals who donated their brains for research.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fneur.2020.570424/full#supplementary-material>



9. Salk JJ, Schmitt MW, Loeb LA. Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. *Nat Rev Genet.* (2018) 19:269–85. doi: 10.1038/nrg.2017.117
10. Xiromerisiou G, Houlden H, Sailer A, Silveira-Moriyama L, Hardy J, Lees AJ. Identical twins with Leucine rich repeat kinase type 2 mutations discordant for Parkinson's disease. *Mov Disord.* (2012) 27:1323–4. doi: 10.1002/mds.24924
11. Schneider SA, Johnson MR. Monozygotic twins with LRRK2 mutations: genetically identical but phenotypically discordant. *Mov Disord.* (2012) 27:1203–4. doi: 10.1002/mds.24991
12. Nacheva E, Mokretar K, Soenmez A, Pittman AM, Grace C, Valli R, et al. DNA isolation protocol effects on nuclear DNA analysis by microarrays, droplet digital PCR, and whole genome sequencing, and on mitochondrial DNA copy number estimation. *PLoS ONE.* (2017) 12:e0180467. doi: 10.1371/journal.pone.0180467
13. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* (2007) 35:W71–4. doi: 10.1093/nar/gkm306
14. Sherry ST. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* (2001) 29:308–11. doi: 10.1093/nar/29.1.308
15. Lodato MA, Woodworth MB, Lee S, Evrony GD, Mehta BK, Karger A, et al. Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science.* (2015) 350:94–8. doi: 10.1126/science.aab1785
16. Proukakis C, Houlden H, Schapira AH. Somatic alpha-synuclein mutations in Parkinson's disease: hypothesis and preliminary data. *Mov Disord.* (2013) 28:705–12. doi: 10.1002/mds.25502
17. Nicolas G, Veltman JA. The role of de novo mutations in adult-onset neurodegenerative disorders. *Acta Neuropathol.* (2019) 137:183–207. doi: 10.1007/s00401-018-1939-3
18. Abyzov A, Tomasini L, Zhou B, Vasmatzis N, Coppola G, Amenduni M, et al. One thousand somatic SNVs per skin fibroblast cell set baseline of mosaic mutational load with patterns that suggest proliferative origin. *Genome Res.* (2017) 27:512–23. doi: 10.1101/gr.215517.116
19. Keogh MJ, Wei W, Aryaman J, Walker L, van den Aamele J, Coxhead J, et al. High prevalence of focal and multi-focal somatic genetic variants in the human brain. *Nat Commun.* (2018) 9:4257. doi: 10.1038/s41467-018-06331-w
20. Hirsch P, Tang R, Abermil N, Flandrin P, Moatti H, Favale F, et al. Precision and prognostic value of clone-specific minimal residual disease in acute myeloid leukemia. *Haematologica.* (2017) 102:1227–37. doi: 10.3324/haematol.2016.159681
21. de Kock L, Wang YC, Revil T, Badescu D, Rivera B, Sabbaghian N, et al. High-sensitivity sequencing reveals multi-organ somatic mosaicism causing DICER1 syndrome. *J Med Genet.* (2016) 53:43–52. doi: 10.1136/jmedgenet-2015-103428
22. de Leng WWJ, Gadellaa-van Hooijdonk CG, Barendregt-Smouter FAS, Koudijs MJ, Nijman I, Hinrichs JWJ, et al. Targeted next generation sequencing as a reliable diagnostic assay for the detection of somatic mutations in tumours using minimal DNA amounts from formalin fixed paraffin embedded material. *PLoS ONE.* (2016) 11:e0149405. doi: 10.1371/journal.pone.0149405
23. Beyens M, Boeckx N, Van Camp G, Op de Beeck K, Vandeweyer G. pyAmpli: an amplicon-based variant filter pipeline for targeted resequencing data. *BMC Bioinform.* (2017) 18:554. doi: 10.1186/s12859-017-1985-1
24. Nishioka M, Bundo M, Ueda J, Katsuoka F, Sato Y, Kuroki Y, et al. Identification of somatic mutations in postmortem human brains by whole genome sequencing and their implications for psychiatric disorders. *Psychiatry Clin Neurosci.* (2018) 72:280–94. doi: 10.1111/pcn.12632
25. Deshpande A, Lang W, McDowell T, Sivakumar S, Zhang J, Wang J, et al. Strategies for identification of somatic variants using the ion torrent deep targeted sequencing platform. *BMC Bioinform.* (2018) 19:5. doi: 10.1186/s12859-017-1991-3
26. Barnell EK, Ronning P, Campbell KM, Krysiak K, Ainscough BJ, Sheta LM, et al. Standard operating procedure for somatic variant refinement of sequencing data with paired tumor and normal samples. *Genet Med.* (2019) 21:972–81. doi: 10.1038/s41436-018-0278-z
27. Sepe S, Milanese C, Gabriels S, Derks KWJ, Payan-Gomez C, van IJcken WFJ, et al. Inefficient DNA repair is an aging-related modifier of Parkinson's disease. *Cell Rep.* (2016) 15:1866–75. doi: 10.1016/j.celrep.2016.04.071
28. Sala Frigerio C, Lau P, Troakes C, Deramecourt V, Gele P, Van Loo P, et al. On the identification of low allele frequency mosaic mutations in the brains of Alzheimer's disease patients. *Alzheimer's Dement.* (2015) 11:1265–76. doi: 10.1016/j.jalz.2015.02.007
29. Nicolas G, Acuña-Hidalgo R, Keogh MJ, Quenez O, Steehouwer M, Lelieveld S, et al. Somatic variants in autosomal dominant genes are a rare cause of sporadic Alzheimer's disease. *Alzheimer's Dement.* (2018) 14:1632–9. doi: 10.1016/j.jalz.2018.06.3056
30. Lee M-H, Siddoway B, Kaeser GE, Segota I, Rivera R, Romanow WJ, et al. Somatic APP gene recombination in Alzheimer's disease and normal neurons. *Nature.* (2018) 563:639–45. doi: 10.1038/s41586-018-0718-6
31. Park JS, Lee J, Jung ES, Kim MH, Kim I Bin, Son H, et al. Brain somatic mutations observed in Alzheimer's disease associated with aging and dysregulation of tau phosphorylation. *Nat Commun.* (2019) 10:3090. doi: 10.1038/s41467-019-11000-7
32. Mokretar K, Pease D, Taanman J-W, Soenmez A, Ejaz A, Lashley T, et al. Somatic copy number gains of  $\alpha$ -synuclein (SNCA) in Parkinson's disease and multiple system atrophy brains. *Brain.* (2018) 141:2419–31. doi: 10.1093/brain/awy157
33. Perez-Rodriguez D, Kalyva M, Leija-Salazar M, Lashley T, Tarabichi M, Chelban V, et al. Investigation of somatic CNVs in brains of synucleinopathy cases using targeted SNCA analysis and single cell sequencing. *Acta Neuropathol Commun.* (2019) 7:219. doi: 10.1186/s40478-019-0873-5

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Leija-Salazar, Pittman, Mokretar, Morris, Schapira and Proukakis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.