Author’s response to reviews

Title: Targeted next-generation sequencing identifies novel variants in candidate genes for Parkinson’s disease in Black South African and Nigerian patients

Authors:

Oluwafemi G. Oluwole (oluwafemi.oluwole@uct.ac.za)
Helena Kuivaniemi (hkuivaniemi@sun.ac.za)
Shameemah Abrahams (shameemah@sun.ac.za)
William L. Haylett (whaylett@sun.ac.za)
Alvera A. Vorster (vorster@sun.ac.za)
Carel J. van Heerden (cjvh@sun.ac.za)
Colin P. Kenyon (kenyon@sun.ac.za)
David L. Tabb (dtabb@sun.ac.za)
Michael B. Fawala (bimbofawale@live.com)
Taofiki A. Sunmonu (taosunmonu@yahoo.com)
Abiodun Ajose (abiodun_ajose57@yahoo.com)
Matthew O. Olaogun (mobolaogun@yahoo.co.uk)
Anastasia C. Rossouw (stacy.rossouw@gmail.com)
Ludo S. Hillegondsberg (ludovhill@gmail.com)
Jonathan Carr (jcarr@sun.ac.za)
Owen A. Ross (Ross.Owen@mayo.edu)
Morenikeji A. Komolafe (adeyoyin2001@yahoo.com)
Gerard Tromp (gctromp@sun.ac.za)
Soraya Bardien (sbardien@sun.ac.za)

Version: 1 Date: 29 Jul 2019
Author’s response to reviews:

Response to reviewers’ comments

29 July 2019

To: Manuscript Editor: Anne Menard

BMC Medical Genetics

Dear Ms Menard

Submission of revised manuscript ‘Targeted next-generation sequencing identifies novel variants in candidate genes for Parkinson’s disease in Black South African and Nigerian patients’ (Ms. Ref.: MGTC-D-19-00107)

We thank the reviewers for their thorough review and constructive comments, which have greatly improved the quality of our manuscript. Thank you for providing us an opportunity to submit a revised version. Our responses to the questions and comments are provided below.

Furthermore, as requested, the manuscript has been significantly shortened:

1. The main text (Background, Methods, Results and Discussion) was shortened from 31.5 to 12.5 pages.
2. The Abstract was shortened from 349 to 285 words.
3. The protein-protein interaction network analyses were removed.
4. The text and the Venn diagram related to the shared variants (Figure 2) were removed.
5. The Radar plots (Figure 3) and the protein modelling figures (Figures 4 - 6) have been moved to Supplementary data.
6. The number of Supplementary files was reduced from 18 to 12.
REVIEWER #1

This work is interesting both for the clinical-genetic aspect of PD and for the methods and analysis shown. It is friendly and easy to read and should be published as it is an interesting contribution to the research of PD in the African population.

The following are some minor suggestions and questions I would like the authors to do.

1. All coding regions of the genes targeted in the panel are meant to be covered. My question is about the exonic flanking region. How many bases are targeted in the design?

RESPONSE: The Ion AmpliSeq™ Neurological Research Custom Panel included a total of 1,619,178 bases in intronic regions linked to exons. The details of these regions are in the “intron-regions_included.bed” file provided as supplemental information. We have inserted the following sentence into the Methods section on page 7, lines 123 - 124, “The intronic regions incorporated as part of the exon targets are listed separately in Additional file 3: Table S3.”

2. The link to the BMC website of figure 1, to the interactive html-version, needs to be reviewed as it is not working.

RESPONSE: Thank you for informing us about this problem. We will ask the BMC Editorial Staff to correct this problem.

3. It is suggested in the discussion that common mutations could have not been detected because of a probable low quality of these loci. It would be interesting to show these data and to independently genotype these variants by other methods. Even if the main objective of the study seems to identify putative rare variants, there is much discussion and work around more common mutations and therefore, these will help to reinforce both results and discussion.

RESPONSE: The reviewer makes a good point. To address the reviewer’s concern we have catalogued all the variants in the known PD genes on the panel, which passed the initial QC, without applying any of the strict filtering criteria. These variants are now listed in Additional file 8: Table S5. Based on this analysis it is clear that none of the patients harboured the common mutations in the genes screened here. We have inserted this sentence into the Results section on page 11, lines 217 - 219: “In addition, we separately screened 16 PD genes on the panel (SNCA, LRRK2, PRKN, PINK1, PARK7, ATP13A2, EIF4G1, GIGYF2, PLA2G6, FBXO7, VPS35, MAPT, HTRA2, SPG11, GRN and DCTN1) for all sequence variants and these results are shown in Additional file 8: Table S5).“.
Moreover, we are in the process of designing and optimising a custom-panel for screening all of the currently known PD genes and these study participants will be screened again using this panel.

4. The commercially available gene panel is mentioned as a limitation for the lack 18 genes previously identified to be related to PD. This is ended something that could have been solved designing a custom panel. Actually, the authors should make clear if analyzing these 34 PD related genes is more important than analyzing the rest of the genes included in the commercial "Neurological Research" panel. Other limitations like coverage and read depth could have been avoided this way.

RESPONSE: We agree that the use of a targeted panel that contains all known PD genes would be important in future studies and, as stated above, we are currently in the process of designing a panel for screening of the known PD genes. This is mentioned in the Discussion on page 16, line 331 - 332: “Future studies will involve using a custom-panel that captures all of the known PD genes ……”

However, the use of the Neurological Research panel for this study was a strategic decision taken at the time, as it was hypothesized that it would be beneficial to screen not only the known PD genes but also candidate genes involved in other neurological disorders and in brain function. Our previous work on South African patients have not identified many of the known mutations and other previous studies on Nigerian patients have not yielded many mutations which is why we decided to use a panel to screen a large number of genes to possibly identify novel candidates for PD in our study participants.

5. Another limitation of the study is mentioned to be the lack of analysis for the variants in ethnicity-matched control samples. In the absence of previous studies, an analysis of the rare variants found in this study should have been performed. And mostly for the ones not found in the PDmutDB database. Previous to functional studies, a substantial number of ethnicity-matched control chromosomes should be genotyped.

RESPONSE: The reviewer makes an important point. We have partially addressed this comment by determining the minor allele frequency for the 60 prioritised variants in the GnomAD database (https://gnomad.broadinstitute.org/) and in the subset of African and African American controls (n =8,128). This information has been inserted into Table 2.
In the future, we will recruit and screen ethnic-matched controls from the South African and Nigerian population to determine the frequency of the prioritised variants. This is mentioned in the Discussion on page 16, line 333 – 334, as part of future studies: “ … and recruitment of a large number of ethnic-matched controls to determine the frequency of prioritized variants in these populations. “

REVIEWER #2

1. Line 149 More detail description of patients is necessary. "Familial" and "sporadic" - what is familial? Clear mendelian inheritance or familial history of PD?

RESPONSE: Two of the South African patients (s43_59 and s94_69) have a clear Mendelian inheritance in their family with both having an affected sibling and an affected parent). This information has been inserted into the Results section (page 10, lines 206 - 207).

We have revised the text to read: “Two of the South African PD patients had a positive family history with both having an affected sibling and an affected parent.”

2. Table 2 - It is necessary to translate the coding of mutations into a common format. In the presented data is difficul to analyze. And add data on population frequencies and estimates from MetaLR and MetaSVM. ?

RESPONSE: As requested, we have edited the naming of the variants to a standard format in Table 2 e.g. Y535F.

In addition, we have added to Table 2, data on the population frequencies of these variants from the GnomAD database as well as the rs numbers from dbSNP. Furthermore, we have included in this Table the pathogenicity scores from MetaLR and MetaSVM.

3. Table 3 - Many possible mutations in one patient. You must to solve a problem of most significant from them and discuss this situation.

RESPONSE: We agree with the Reviewer that it is very unlikely to have many possible mutations in one PD patient. We have therefore removed all reference to this in the text but have left the results in Table 3 as these may be useful for future studies when deciding whether the variants are pathogenic mutations or common polymorphisms.
4. Figure 1 - repeats the text of the article in an unsuccessful form. It must be removed.

RESPONSE: Figure 1 is a Sunburst diagram showing the functional classes of the 14,655 rare sequence variants identified in the patients. Unfortunately, it seems that the interactive HTML-version uploaded during the submission is not working properly. We will ask the BMC Editorial Staff to correct this problem and kindly request that the reviewer considers our request to keep this comprehensive summary of all categories of sequence variants identified in this study, which may be useful to other researchers.

5. Figure 2 is also an unsuccessful and poorly readable form. This is about comparing samples from SA and Nigeria - and not about PD. It is also better to remove, leaving a comparison in the text.

RESPONSE: We agree that this analysis is more about comparing the two populations than it is about PD, and have therefore removed this figure (Venn diagram) and all of the related text from the manuscript.

6. Figure 3 and Line 297 - "Blind" drawing. Replace with a table with cut-off values for falling into the "deleterious" category

RESPONSE: Figure 3 provided radar plots to illustrate 17 scores for the 60 variants considered deleterious. These plots also show the concordance between the algorithms used. We consider this to be an important finding, but since it may not be directly relevant for this study, we have moved this figure to Additional file 9: Figure S3. We hope that the reviewer finds this acceptable.

7. Line 168, 211 - cut the whole section by removing the standard methods

RESPONSE: This section in the Methods has now been removed from the main document and inserted into Supplementary data.

We have added the sentence on page 7, lines 124 – 125, “Details on the library construction and next-generation-sequencing protocols are available in Additional file 4.”

8. Line 294 - Why do you use different values for cutting in the case of known genes PD and new genes? These parameters relate to mutations - not genes. And what are your cut-off parameters in the further actively discussed mutations in the genes of familial forms of PD?
RESPONSE: The reviewer has made an important point and we have now decided to remove this distinction from the data analysis. However, since it is important to capture all sequence variation in the established PD genes such as PRKN and PINK1 in these understudied African populations, we have decided to include an Additional file 8: Table S5, which provides a catalogue of this data. For these genes, we have not used any stringent cut-off parameters but instead have catalogued all variants that passed the initial QC.

9. Line 319 - remove the entire section on modeling. You do not model a protein with a mutation - you just show the location of its localization on the model. The location of a mutation in a particular functional domain of a protein can simply be said in the text.

RESPONSE: We kindly request that the reviewer allows to keep the modelling information in the manuscript. In the absence of wet-lab functional studies, it is necessary to determine the possible effect of the missense variant on the protein using in-silico protein modelling. We acknowledge that the findings have their limitations based on the model used but these results are usually the starting point to decide whether or not future functional studies should be done. However, we have moved all of the modelling figures out of the main document and into Supplementary data i.e. Additional files 10 – 12: Figures S4 – S6.

10. Line 416, frameshift mutations. Further, nothing is said about these mutations - although they may be pathogenetically highly significant.

RESPONSE: It is acknowledged that frameshift mutations are important as they account for a significant proportion of the pathogenic mutations that lead to human diseases. However, since they usually occur in repeat regions of the DNA, and the platform (IonTorrent) used here is known to not be as reliable in detecting these types of mutations, we therefore, concentrated our efforts on characterizing the single nucleotide variants. However, we have provided the list of frameshift variants found for future validation studies.

11. Line 447, "We focused on identifying rare (MAF ≤ 0.01) or novel" - but some your possibly deleterious variants are very frequent (Line 489-497). A population frequency of this variants?

RESPONSE: When shortening the manuscript it was decided that this text should be removed from the Results section.
12. GENERAL REMARK - the article should be radically reduced with a clear emphasis on key results.

RESPONSE: This comment has been addressed, and the main text of the manuscript has been reduced from 31.5 to 12.5 pages. In addition, Figure 2 and accompanying text has been removed and the protein-protein interaction network analysis has been removed, while Figures 3 – 6 have been moved to the Supplementary material.

REVIEWER #3

The manuscript describes the results of a study sequencing 751 genes in a total of 57 patients with Parkinson's disease (PD). 33 of the patients were Black South African and 14 were Nigerian. The sequencing panel was a neurological research panel and included the most important genes found in familial PD. The authors report that they found more than 14000 rare variants in these patients, including 60 rare variants in 44 genes that were predicted to be deleterious of which 7 rare variants are in three known PD genes.

In most populations, PD is generally a sporadic disease and pathogenic mutations in Mendelian genes are found in 2-5 % of patients. There are very few studies from Africa, so the contributions of known genes are largely unknown. Studies from such populations are warranted. Despite this there are in my opinion a number of major limitations of the current study:

1. It is unclear to me what the aim of the study actually is. The sample size is very small for a genetic study, and the only thing that can be done is to examine known disease related genes for mutations. However, the patients are not enriched for familial disease or young onset, limiting the chance of finding mutations in known PD genes. There are no families to find new disease genes, no power to do association studies, and no controls to study the frequency of variants found in PD patients in the general population. As a consequence of this, it is very limited which conclusions you can draw based on the presented data.
We apologise if the overall aim of the study was not clear and we thank the reviewer for the pertinent and important points raised. The primary goal of the study was to determine whether there was a common pathogenic mutation in this series of sub-Saharan African (SSA) PD patients from South African and Nigeria, e.g. similar to the ~40% frequencies observed for LRRK2 G2019S in North African. A secondary goal was to characterize the genetic variation in known and novel genes for PD. We therefore used an NGS-gene panel approach to capture the sequence variation across 751 candidate genes related to neurological disorders including LRRK2, PARK7, PINK1, PRKN, SNCA and ATP13A2. Additionally, another aim of this study is to raise awareness for the need of larger sample collections for genetic studies, where we can then select specific samples for exome/whole-genome sequencing. We did not observe a common ‘pathogenic’ mutation across the samples and this reflects that the genetics of PD in SSA populations will likely be heterogenic as observed across white Europeans and Asian populations.

However, we did detect a number of rare, novel variants and, although it is acknowledged that the sample size is limited, we believe these findings do provide a foundation for future studies on PD patients from the African continent. We have now provided the full sequence variation across 16 PD genes on the panel in Additional file 8: Table S5, which may be of use to other researchers. Additionally, we have included the GnomAD ‘control’ frequencies for each variant in Table 2. This is noted for all populations, and separately for the African populations, so as to highlight ethnic-specific variants and allele frequencies that may also help inform studies in other populations, e.g. haplotype structure studies that could help resolve functional variants under the large European GWAS peaks. We appreciate the many limitations of the study but hope the reviewer will appreciate the efforts and wealth of data that has been generated. With the data made available to other researchers, we hope this will encourage other centres in SSA to perform similar studies and meta-analyses to answer some of the questions the reviewer has raised.

2. The manuscript is extremely long and detailed, both the abstract and the main text. If the data should be presented as a scientific publication, it needs to be much more stringent in its form.

RESPONSE: This concern has now been addressed and the main text of the manuscript has been significantly shortened from 31.5 to 12.5 pages. Figure 2 and accompanying text has been removed and the protein-protein interaction network analysis has been removed while Figures 3 – 6 have been moved to the Supplementary material.
3. The authors present that the patents had a total of more than 14000 rare variants in the 751 genes, is this more than expected? How to determine which variants that may have a role in disease?

RESPONSE: It is difficult to provide an answer to this question since these populations have not been well characterised from a genetic perspective. Therefore, whether the 14,000 variants identified is more or less than expected, is not currently known. Initially, determining the frequency of these variants in ethnic-matched controls and in the future, functional studies are needed to determine whether the variants play a role in PD.

4. On page 13 it is stated that "In order to identify novel PD candidate genes..." How should this be possible given the study design?

RESPONSE: On page 8 (lines 153 - 154), these words have been deleted and the sentence now reads “We focused on variants that are rare in control populations as defined by a minor allele frequency (MAF) threshold of 0.01.”

5. For the study of rare variants in known PD related genes a less stringent cut off for predicted deleteriousness was chosen. Is there any precedence for this and why was 0.45 selected as a cut-off?

RESPONSE: This concern was also stated by reviewer #2 (comment no. 8), and we have now addressed this by removing the cut-off of 0.45 for the known PD genes. Instead, in order to capture all sequence variation in these genes we have provided a list of the variants identified in Additional file 8: Table S5.

6. Based on the previously mentioned limitations to the study design it makes little sense to perform network analyses based on the results.

RESPONSE: As requested, the protein-protein interaction network analyses have been removed from the manuscript.