Author’s response to reviews

Title: Lessons learned from Whole Exome Sequencing reanalysis: Our experience on 200 Lebanese exomes

Authors:

Nadine Jalkh (nadine.jalkh@usj.edu.lb)
Sandra Corbani (sandra.corbani@usj.edu.lb)
Zahraa Haidar (zahraa.haidar@net.usj.edu.lb)
Nadine Hamdan (nadine.hamdan@net.usj.edu.lb)
Elias Farah (elias.farah@usj.edu.lb)
Joelle Abou Ghoch (joelle.aboughoch@usj.edu.lb)
Rouba Ghosn (rouba.ghosn@usj.edu.lb)
Nabiha Salem (nabiha.salem@usj.edu.lb)
Ali Fawaz (dralifawaz@hotmail.com)
Claudia Djambas Khayat (claudiakhayat@yahoo.fr)
Mariam Rajab (drmariam1@hotmail.com)
Chebl Mourani (cheblmourani@gmail.com)
Adib Moukarzel (amoukarzel@pol.net)
Simon Rassi (srassi@hotmail.com)
Bernard Gerbaka (bernard.gerbaka@usj.edu.lb)
Hicham Mansour (hicham.mansour@gmail.com)
Malek Baassiri (malek-baassiri-med@hotmail.com)
Rawane Dagher (rawanedagher@gmail.com)
David Breich (drdavidbr@hotmail.com)
André Mégarbané (andre.megarbane@yahoo.fr)
November, 5th 2018

Dear Editor,

The authors would like to thank both reviewers for their helpful comments, corrections and suggestions.

Please find below the List of Changes to:

Reviewer 1

Q1- The patients’ diagnoses success rate for the diseases they studied lacked a statistical interpretation based on the population sub-strata. IS there such interpretation? Am I missing something?

R1- Each identified mutation was compared to different databases: 1000Genome, ExAC and gnomAD, that include data from more than 100 000 samples. In addition to that, the detected mutations were checked in our local Lebanese WES database that includes data from 330 individuals.

In WES, this is the only possible statistical population analysis.

Q2- Whilst reanalyzing the exomes, what were the underlying outliers from previous analysis?

R2- First WES analysis yielded a molecular diagnosis in 49.5% of the cases.
The reanalysis allowed us to increase this percentage to a yield of 56%. The outliers from the first analysis (the cases that were undiagnosed in the first round but were able to benefit from a molecular diagnosis after the second round of analysis), are listed in Table 3 of the manuscript.

Q3- The point on technical limitation in page 7 needs to be revisited. It is expected with tools such as CGH but isn't it more nice to go for integrated (exome+transcriptome) instead?

R3- This study aimed to perform molecular diagnostics for patients with genetic disorders, by WES. The discussion in the manuscript was mainly focused on the advantages and disadvantages of WES.

We agree with the reviewer that, in cases of negative results with WES, transcriptome is one of other techniques that may be applied to complete the molecular investigation.

In fact, to be able to reveal genetic defects that are missed by WES, we may need to perform:

CGH array to detect large duplications/deletions, transcriptome to reveal any defect in gene expression and ideally Whole Genome Sequencing to uncover intronic variations affecting gene expression and big translocations, deletions and insertions that would alter a gene.

However, the current trend in human genetic studies is to start with WES and CGH.

Q4- On the point where there must be communication between the geneticists and clinicians, it would be nice if the authors mentioned about the precision medicine.

R4- We agree with the reviewer about the importance of mentioning “the precision medicine” especially that the molecular diagnosis nowadays is allowing physicians to prescribe the targeted treatment in many cases. This idea was mentioned in the conclusion part “…all driving towards personalized medicine”.

Q5- The sfari.org URL cited in page 8 could have an accurate URL. If this is rendered, then please add "last accessed date"

R5- The URL cited in page 8 was replaced by the last database updated on August 13, 2018.

Q6- The authors in conclusion make a note that the WGS could perhaps be performed to detect ncRNAs. I don't think the depth would be that efficient to get the ncRNAs and even if they do, they might end up with mere identifying them, not characterize them for bona fide. Instead, it would have been nice to extend the targets from WES to intergenic regions or go for transcriptome analyses to leverage better understanding of ncRNAs. This must be elaborated. I would like to suggest authors to get articles in PubMed on "Identifying ncRNAs from exomes" It is still in infancy, though. A point on it would be nice.

R6- The reviewer is right that the field of research on ncRNAs is still in development. There might be a link between ncRNAs and some genetic diseases. However, these data can not be used for diagnostics purposes, yet.

In our conclusion, we mentioned the use of WGS for the identification of non–coding mutations but we meant all missense mutations that are in intronic regions that could affect the expression of coding mRNAs.

Q7- in Page 10, while the authors focused on intron-exon boundaries, a careful assessment (BLAST etc.) Would entail identifying ncRNAs. Did the authors find any

R7- WES covers around 20 bp from both sides of an exon. These areas were checked carefully for the identification of splicing mutations present in non-coding regions.

Q8- On a hindsight, the subheads mentioned in italics could be in small sentences.

R8- The subheads mentioned in the discussion part were formatted in italics to attract the attention of the reader.

Q9- A pictographic flowchart of methods would be nice. The tables are made more descriptive though.

R9- A reference with a corresponding pictographic flowchart for the method was added in the Methods section of the manuscript.
Reviewer 2

Abstract

Q1- “connecting a rare variation”. Should be “connecting a rare genetic variation”

R1- As suggested by the reviewer, the expression “connecting a rare variation” was corrected in the abstract (changes are tracked in the word document).

Q2- “a clear phenotype with a detailed family history and ending, in some cases, with functional assays that are crucial for the validation of the pathogenicity of a mutation”. From the abstract and this part in particular, it is not very clear that the authors are actually aiming to identify the genetic cause of specific diseases. I would suggest rewriting the first part of the abstract accordingly.

R2- This sentence was modified in the abstract as suggested by the reviewer (in Track changes).

Q3- “a wide spectrum of genetic disorders”. I would recommend listing at least some of them for clarity.

R3- As suggested by the reviewer, listing some of the studied genetic disorders was added to the abstract.

Q4- “49.5% overall success rate”. The meaning of “success rate” should be explained/described (i.e. the specific genetic mutation causing the phenotype/disease)

R4- As suggested by the reviewer, the expression “49.5% overall success rate” was elaborated in the abstract.

Q5- “diagnostic yield”, not clear what this really means (sensitivity?)

R5- Diagnostic yield refers to the success in molecular diagnostics, in other word to the percentage of cases where the disease causing mutation is identified. This term is usually used in Next Generation Sequencing analysis.

Q6- “were reanalyzed”, how? Please briefly describe what was changed.
R6- In response to the reviewer’s comment, this statement was further explained and developed in the main text (Results section- Track changes).

Main text

Q7- “An initial analysis was performed, followed by data re-analysis”, should be described in more detail.

R7- In response to the reviewer’s comment, this statement was further explained and developed in the main text (Results section- Track changes).

Q8 “The introduction should be expanded by describing the importance of having a genetic test for the certain diseases, methods currently available and the relevance of their study within the context”.

R8- We would like to thank the reviewer for his input. As recommended, we further elaborated the importance of having a genetic test in the introduction section of the manuscript (Track Changes).

Q9- “All types of mutations (small frameshift, nonsense, splice sites, and missense mutations) were identified.” Suggest rewriting.

R9- In response to the reviewer’s comment, this sentence was re-written in the manuscript (Track changes).

Q10- “all “private” Lebanese polymorphisms”. Not clear.

R10- The term “private” is used to indicate genetic variations that are specific for a population.

Q11- The list of “lessons learned” in the discussion is mix of a bit of everything. I would rather move the description of the specific mutations identified to the results.

R11- All the specific mutations mentioned in these paragraphs are already stated in tables 2 and 3 in the Result section of the manuscript. They were elaborated in the discussion to support the listed ideas.
Q12- “of low-cost molecular diagnostics”. Is WES low cost? I believe the authors refer to other tools, this therefore should be clarified.

R12- The gold standard of genetic diagnosis was direct sequencing also known as Sanger sequencing. This will remain a low-cost technique when the gene size is small. However, in the case of a large gene (mainly large exons, such as the CFTR gene ...) or heterogeneous disease where more than hundred genes are involved, WES is considered a “low-cost” technique. All patients included in this study presented with heterogeneous diseases, thus justifying the use of WES.

Q13: A few English grammar/typos to be corrected along the text.

R13: The manuscript was carefully read and corrections were made.