Author’s response to reviews

Title: Targeted next generation sequencing as a tool for precision medicine

Authors:

Markus Gulilat (bgulilat@uwo.ca)
Tyler Lamb (tyler.lamb@ucalgary.ca)
Wendy Teft (wendyteft@gmail.com)
Jian Wang (jwang@robarts.ca)
Jacqueline Dron (jdron@uwo.ca)
John Robinson (robinson@robarts.ca)
Rommel Tirona (rommel.tirona@schulich.uwo.ca)
Robert Hegele (hegele@robarts.ca)
Richard Kim (richard.kim@lhsc.on.ca)
Ute Schwarz (ute.schwarz@lhsc.on.ca)

Version: 2 Date: 12 Apr 2019

Author’s response to reviews:

Dr. Matteo Pasini
Editor
BMC Medical Genomics

Re: Manuscript Resubmission to BMC Medical Genomics – MGNM-D-19-00009R1

Title: Targeted next generation sequencing as a tool for precision medicine
April 12, 2019

Dear Dr. Pasini,

Thank you for your response concerning our previously submitted manuscript entitled “Targeted next generation sequencing as a tool for precision medicine”, and the opportunity to resubmit a revised manuscript which takes into account the reviewers comments.

We have edited the manuscript to address the reviewer's comments, and provide a detailed point-by-point response in the Comments section below.

We believe that the successful application of a targeted NGS gene panel to assess novel variation in pharmacogenes and CYP2D6 copy number is of interest to the field of precision medicine and the readers of this Journal. Accordingly, we would like to ask to reconsider our revised manuscript for publication in BMC Medical Genomics.

All authors have read and approved the revised manuscript for re-submission. There is a US provisional patent application for the PGxSeq panel pending (RB Kim, RG Tirona, UI Schwarz), while other coauthors have no conflicts of interest. There are no issues relating to journal policies. The work is original, has not been previously published, and is not being considered for publication elsewhere, in whole or in part, in any language.

We appreciate your time and look forward to your response.

Sincerely,

Ute Schwarz
Assistant Professor
RESPONSE TO COMMENTS OF THE REVIEWERS

We would like to thank the reviewers for thoughtful review of the manuscript. Please find a detailed point-by-point response below concerning the specific comments and concerns. All changes in the revised manuscript have been highlighted.

Reviewer #1:

1. The authors do not discuss the limitations of their study, in the sense that it is not able to detect pharmacogenomic variants in the introns and elsewhere.

Response: We agree with the reviewer that, compared to whole genome sequencing (WGS), our targeted exome panel is unable to detect pharmacogenomic variants in 3’- and 5’-untranslated and intronic regions with the exception of those that were specifically targeted in our panel. While the discovery of such variants is of similar importance, the majority of currently available prediction algorithms concerning functional implications (i.e. SIFT, PolyPhen, Mutation Taster) are limited to the assessment of coding variants, representing a major obstacle for the evaluation of non-coding variants. Consequently, these markers remain of uncertain significance without further in vitro or in vivo testing. Despite its limitation to coding regions, the time- and cost-effectiveness of sequencing, data processing and storage, as well as variant analysis represent primary incentives for a targeted exome approach compared to WGS.

We have now addressed this as a limitation in our Discussion (page 21, lines 456-458): “Lastly, in contrast to whole-genome sequencing, our targeted exome panel is unable to detect
pharmacogenomic variants in 3’- and 5’-untranslated as well as intronic regions that may be of relevance.”.

2. The authors did not compare their findings with similar studies that have been performed previously where whole genome sequencing has been used (e.g. Mizzi et al., 2014 and references therein).

Response: We thank the reviewer for this comment, highlighting some important research in this area that had not been adequately addressed in our manuscript. We have now included previously reported studies (Mizzi et al. Pharmacogenomics 2014; Ashley et al. Lancet 2010) utilizing whole-genome sequencing for pharmacogenomic profiling in our Background section (citations page 4, line 73), and discuss their findings in comparison to our results. Accordingly, we now detail in the Discussion (page 17, lines 355-358): “Importantly, 60% of the observed SNVs were rare (536 variants; 2.3 per patient) or novel (105 variants; 4.4 per 10 patients), the latter absent in more than 60,000 individuals [45]; a similar frequency of 73% has been previously reported in a whole-genome sequencing study of 231 pharmacogenes [20].”.

3. It would be useful to discuss about the feasibility and cost-effectiveness of the proposed method as compared to panel-based approaches, also deducted by the number of rare and novel variants identified using this novel method. Why someone to use this novel method and not an existing panel-based test?

Response: We would like to thank the reviewer for this helpful comment to compare feasibility/cost effectiveness of existing array-based genotyping assays with our sequencing panel in the context of detecting rare or novel variation. While genotyping arrays may be more cost effective, such arrays will not identify novel variation, which represents a significant portion of SNVs in pharmacogenes as previous and our research indicates. Moreover, the capacity of assessing CNV in pharmacogenes is another advantage of our NGS panel compared to available array-type methods. We have now extended our Discussion as follows (page 17, lines 359-363): “… Accordingly, a significant portion of novel variation will likely be missed when utilizing more cost-effective, array-based genotyping platforms such as DMET+ (1936 SNVs in 231 pharmacogenes; Affymetrix, CA, USA) or the genome-wide Infinium Global Screening Array-
24 (665,608 SNVs; Illumina, CA, USA). Moreover, the capacity of assessing CNV in pharmacogenes is an additional advantage of this NGS panel.”.

4. Also, the authors only vaguely responded to the first round of comments, which does not allow to assess the extent of the review performed in this revision round.

Response: Previously, the authors had only corresponded with the Senior Assistant Editor via e-mail. We had been asked to correct a minor formatting issue, and addressed a question concerning the pediatric/ parental consent (as noted under Declarations, Section Ethics approval and consent to participate). While a revised manuscript was uploaded, no formal response letter had been requested at that time.

Reviewer #2:

1. When listing concordance values, please include the confidence interval as well.

Response: We thank the reviewer for this comment, and have now added the confidence interval for sensitivity and specificity in our Results (page12-13, lines 259-262): “While we did not detect any false positive results (a variant was detected by NGS but not confirmed by TaqMan genotyping; specificity of 100%, 95% CI, 100%-100%), a false negative NGS result was observed in two heterozygous carriers for DPYD rs67376798 and CYP2D6*10 rs1065852 (no variant detected by NGS but observed by TaqMan genotyping; sensitivity of 99.7%, 95% CI, 99.2-100%).”

2. Switching between different sets of reference SNPs seems odd. When performing analysis of sequencing data for 146 pharmacogenes they used ESP and 1K genomes as reference.
Later they used ExAC and dbSNP for population frequencies. Why not use gnomAD at both steps? It is a much larger data set.

Response: We agree with the reviewer that the Genome Aggregation Database (gnomAD) is the most comprehensive variant resource currently available including exome-wide data of more than 125,000 individuals. However, we believe that the ExAC database with exome data of 60,706 individuals represents a suitable alternative. We would like to note that we consistently used the ExAC/ dbSNP databases for the analysis of our sequencing data. The Exome Sequencing Project (ESP) and the 1000 Genomes Project (1000G) the reviewer is referring to was only mentioned in the Introduction of the manuscript in relation to a previous report that assessed 146 pharmacogenes (Kozyra et al. Genet. Med. 2017).

3. The authors mention the need for "careful probe design" to sequence "genes with high sequence homology, nearby pseudogenes and complex structure". The methods presented only indicate using the Illumina Design Studio but don't detail any special steps taken for capturing these regions.

Response: Technical challenges of NGS have been widely discussed for pharmacogenes due to their high sequence homology, nearby pseudogenes or complex structure, as noted in the Introduction. As previous reports and our results show (Gordon et al. Pharmacogen Gen 2016; Han et al. CPT 2016), with a few exceptions, deep coverage and high concordance rates (> 99.8%) can be achieved, and thus support the application of NGS platforms for these genes. Exceptions may include but are not restricted to major histocompatibility complex genes (i.e. HLA-B and HLA-DQB3) and select regions within CYP2D6 (92.9% concordance, Han et al. CPT 2016; 98.1%, this study), CYP2A6, or CES1 (regions with lack of coverage, as observed in this study). While no specific steps had been taken by us at the probe design stage, we carefully assessed depth of coverage across select genes known to be challenging as well as determined variant accuracy.

4. Algorithms such as SIFT, PolyPhen-2, and CADD are much better at predicting loss of function events than gain of function events. The authors did not discuss the expected higher
false negative rate for correctly characterizing rare/novel gain of function variants of clinical interest.

Response: We thank the reviewer for this comment. We agree that previous studies have shown a higher false negative rate for predicting rare/novel gain of function (GOF) variants compared to loss of function (LOF) variants using algorithms such as SIFT and PolyPhen (Flanagan et al. Genet Test Mol Biomarkers 2010), while other algorithms such as CADD may be comparable in their performance for GOF and LOF in case of disease genes (Gosh et al. Genome Biology 2017). We have now modified our Discussion as follows (page 19, line 409-411): “Moreover, a higher false negative rate may apply for predicting rare gain-of-function compared to loss-of-function variants using SIFT and PolyPhen [58], while algorithms such as CADD may be more comparable [59].”