Author’s response to reviews

Title: Protein modeling to assess the pathogenicity of rare variants of SERPINA1 in patients suspected of having Alpha 1 Antitrypsin Deficiency

Authors:

Friedrich Kueppers (Friedrich.Kueppers@tuhs.temple.edu)
Mark Andrake (Mark.Andrake@fccc.edu)
Qifang Xu (Qifang.Xu@fccc.edu)
Roland Dunbrack (rolanddunbrack@gmail.com)
Joannah Kim (joannah_kim@biocerna.com)
Christopher Sanders (csanders@biocerna.com)

Version: 1 Date: 03 May 2019

Author’s response to reviews:

May 03, 2019
Dr Kuivaniemi
Editor, BMC Medical Genetics

RE: Your submission to BMC Medical Genetics - MGTC-D-18-00476

Dear Dr Kuivaniemi

Thank you for considering the above-mentioned paper for publication in BMC Medical Genetics. We have addressed below each of the points raised by the reviewers’ and indicated, when relevant, where changes have been made in the manuscript. All changes made to the text are tracked in the revised manuscript.
We look forward to receiving your feedback and editorial decision.

Yours sincerely,

Friedrich Kueppers

Point-by-point responses to reviewers’ comments

Please view tracked changes ‘inline’ for page/line numbers to correspond

Editor comments:

Major

1) Is this modelling applicable to other diseases?

Response: These predictive methods are applicable to any disease arising from single nucleotide variations; however, their utility in different diseases is likely to vary between different methodologies employed.

2) Abstract (line 37) and Methods (line 190): How was the scoring system PolyPhen-2 selected from about 26 different scoring systems? Is it sufficient to use only one system for comparing the results to the SVM and FoldX?

Response: Most studies report a whole host of scoring systems and since the concordance between them is not that good, it would seem inadequate to rely on one score. How would these results compare to results from Annovar or MetaSVM and MetaLR scores reported by Liu et al. 2016? PolyPhen-2 was selected as it has been used extensively in AATD (e.g., Silva et al., 2016), and has been shown to have good accuracy for this application (Giacopuzzi et al., 2018).
Although it is correct that a large number of SNV prediction methods are available, the purpose of this study was not purely to assess the performance of these measures. The central objective was to raise awareness of the underappreciated contribution of novel/rare variants to the development of lung disease. We therefore felt that presenting a host of prediction data would reduce the clinical/technical balance of this paper. Furthermore, in practice, a clinician would likely not have time to consult a panel of different methodologies before coming to a decision on the best course of treatment.

To address the concern raised here, we have added content to the limitations to acknowledge that there are numerous methods available and each differ in performance for characterizing AAT variants (See discussion, page 22, line 514–524). Additionally, a new column has been added to Table 2 containing the prediction scores of 5 other commonly used methods. They demonstrate a broad agreement with the 3 methods highlighted in this manuscript (SVM, FoldX, and Polyphen-2).

3) Submit sequence data to the European Nucleotide Archive (ENA) and provide the submission number in the methods. Data statement on lines 449-450 is not correct, since the reader would need the sequence data to evaluate and re-analyze the results.

Response: We have previously contacted the editorial office regarding this – we are happy to submit the data to ENA once the manuscript is accepted for publication.

4) Incorporate most of the information from Additional File 1 to the main manuscript. This is information that the readers need to understand the analyses and it should be part of the main manuscript.

Response: All content from Additional File 1, apart from the detailed breakdown of the features of SVM, has been incorporated in the manuscript.

5) Use Suppl. Figure E1 as one of the main figures.

Response: Previous Figure E1 is now Figure 2 in the main text.
6) Highlight limitations of the study such as the small sample size (only 23 patients were assessed).

Response: This has been noted in the limitations section of the discussion (see page 22, line 505–508).

7) Discussion is very long

Response: We feel that a slightly longer discussion is warranted as there are number of topics to cover. Please also note that we have added further points to the discussion in response to comments raised by the editor/reviewers. However, as suggested, we have endeavored to reduce the length of the discussion/conclusions, and it is currently approximately 150 words shorter than the original submission.

Comments – Minor

1) Title: Change “mutations” to “sequence variants” here and throughout the manuscript.

Response: We have made this change throughout; in places we have shortened this to ‘variants.’

2) Title: use lower case first letters for the protein/gene name.

Response: This change has not been incorporated as it does not appear to be the convention with SERPINA1 in the literature/databases, and for other genes studied in BMC Medical Genetics.

3) Title: add the gene symbol to the title

Response: SERPINA1 is the approved gene symbol https://www.genenames.org/data/gene-symbol-report/#!/hgnc_id/HGNC:8941

4) Abstract: indicate number of samples analysed on line 29.

Response: We have added the sample/patient number to the results section of the abstract.
5) Abstract: Change “mutations” to “sequence variants”.
Response: As noted above (response to minor comment 1), this has been amended throughout.

6) Abstract, line 34: insert “a” before “result”
Response: This addition has been made; please note further changes have been made to bring the word count within the journal word limit.

7) Abstract: use SNVs, not SNPs here and throughout the manuscript.
Response: We have made this change throughout

8) Background, lines 53, 54: insert space between a numerical value and its unit.
Response: This change has been made.

9) Background, line 58: change “clinically reported” to “reported to be clinically significant”.
Response: This change has been made.

10) Background, line 76: change “genetic” to “DNA”.
Response: This change has been made.

11) Background, lines 78-82: remove this text. This is a genetics journal and readers are aware of the technique.
Response: As suggested, we have removed most of the background text on NGS (see background, page 5, line 81–88).

12) Line 212: remove “summary”
Response: We have removed ‘summary’ and edited this text (see Methods, Predicting Pathogenicity, page 8, line 159–160).
13) Line 356: what is meant by “structural” mutations?
Response: We have clarified this statement: ‘single amino acid substitutions/deletions leading to subtle structural changes to the AAT protein’ (see Discussion, page 20, line 461–462).

14) Page 28, line 34: insert space before (
Response: This change has been made (see Table, page 34).

15) Supplement states on page 2 “examples of observed IEF bands are shown in Figure 1.”, but no such figure was included in the submission
Response: It was not our intention to submit this Figure as part of the manuscript submission; we have therefore removed mention of it from the methodology.

16) Supplement, page 4: should the name of the program be MUSCLE (all caps)?
Response: Yes – this change has been made (see Additional File 1, page 1).

17) Supplement, page 4: please provide a literature reference for the Naccess program.
Response: We are not aware of a literature citation for this reference. The developers ask for the program to be cited in this manner on the website http://wolf.bms.umist.ac.uk/naccess/
1) In this work, NGS sequencing of the SERPINA1 gene was applied to a US cohort of patients with suspected AATD, based on low AAT plasma levels and on clinical presentation. This analysis led to identify rare SERPINA1 alleles in several cases. The authors then used a set of computational methods to predict the pathogenic potential of the SNPs identified.

Indeed, the discovery of many novel SERPINA1 variations calls for the implementation of novel diagnostic strategies to timely assess their pathological significance. However, the use of predictive algorithms, generally developed on large datasets, may be inconsistent when applied to a specific gene without any experimental validation, and should be critically evaluated.

I suggest a revision of text and figures before the paper is considered for publication.

Here are some suggestions to improve data presentation

Response: We thank the reviewer for their careful review of the paper and suggestions for improvement. Where possible, we have implemented the suggested changes and indicated, when relevant, where changes have been made in the manuscript (please see below). In reference to the reviewer’s point regarding experimental validation of the modeling results, we agree that these computational techniques are not infallible, and we have included a note that experimental validation of these data is required to confirm the pathogenicity of the variants presented (see Discussion, page 22–23, line 519–523).

2) Most of the AAT variations found in this study have been already reported and some are well characterised (eg. In PMID 24084503 (Joly 2014); 29882371 (Giacopuzzi 2018); 29232161 (Matamala 2018); 22330941 (Zhan 2012); 20453271 (Arora 2009); 22008137 (Joly et al 2011); 2696185 (Crystal 1989); 26270547 (Ferrarotti 2015).

Response: We were aware that a number of the variants had already been reported in the literature, including H262Y, V333M, A142D, V210E, M385T (as reported in the footnotes of Table 1). Where applicable, we have added additional references, as highlighted by the reviewer, to cite the above mutations. Variants that were not previously identified in the present paper as being reported in the literature are A325P (Ferrarotti et al., 2015) and I50N (Matamala et al., 2018). We have added these citations and acknowledged the previous reports of these variants. We did not identify any further crossover between our paper and the variants reported by Joly et al., 2011 and Crystal et al., 1989.
3) Computational predictors have been also used and evaluated previously on SERPINA1 variants (eg in 29882371; 27296815). Although the presentation of new cases is certainly helpful to establish an association between such very rare AAT alleles and the disease, reference to previous work should be clear and the new data discussed in view of previous findings.

Response: We have added an additional column to Table 2 to compare scores presented in these papers and the present analysis – it was found that the data generally align. We have added content to the discussion (page 21, line 487–491) on this subject.

4) The AAT levels are better presented along with CRP values, as an inflammatory state could explain normal AAT values in the presence of deficient alleles (eg "There was only one exception from the general agreement between computational predictions and AAT serum levels - the P369H mutation, observed in patient 21034. All three computational analyses predicted the mutation to be highly deleterious; however, the AAT serum level was normal (121.2 mg/dL)".

Response: Where available, CRP data have been added to Table 1; a CRP value was not available for patient 21034. In addition, we have noted in the results that the normal AAT level may have been as a result of an inflammatory state present at the time of sampling (Results, page 13, line 277–278).

5) The IEF results are mentioned in the text but not shown.

Response: As noted above, it was not our intention to include this Figure in the paper, and the associated text has been deleted.
6) The novel variations should be clearly identified, eg by listing them in a separate table, and discussed in details. The outcome of null alleles is pretty clear in the field, but novel amino acid substitutions would benefit from a specific structural evaluation. To my knowledge, novel variants are: I9N (in the leader sequence and therefore unlikely to have an effect; here numbering is not consistent with the conventional nomenclature adopted in this paper, as evident in Figure E1 also for I50N); P28L; Q40R; M221T and A142D. A more focused discussion of these AAT variations, their structural features and their prediction scores, would allow to appreciate the performance and possible flaws of the different predictors used, when confronted with truly unknown cases.

Response: We thank the reviewer for their suggestion; however, we do not feel that an additional table in this case would enhance the manuscript. The structural consequences of all variants are elaborated on in detail in Table 2. As all SNVs are reported in the dbSNP and are therefore not technically ‘novel’, we now describe these variants as ‘rare’ throughout. In response to the reviewer’s suggestion, we have focused more on the variants that, to the best of our knowledge, have not been previously reported in the literature (see Discussion, page 17, line 388–396).

As the reviewer mentions, the current numbering of AAT amino acid changes (without the 24 amino acid precursor) is predominantly how AAT variants are described in the literature, and would make interpretation of the data easier for clinicians/researchers reading the paper. We acknowledge that the description of I9N was not consistent with this numbering convention and to resolve this, we have stated throughout that this variant includes the leader sequence, including in the legend for Figure 2 (previously Figure E1). We thank the reviewer for spotting the inconsistency with regards I50N in Figure E1 (now Figure 2) – we have rectified this in the resubmitted Figure.
Reviewer 2

1) Kueppers et al. presents a well-written report summarizing their evaluation of the utility of computational modeling to provide supporting evidence regarding the pathogenicity of novel SNPs in the SERPINA1 gene. For a total of 23 patients suspected of having Alpha 1 Antitrypsin deficiency they used NGS and predictive computational analyses, in addition to quantification of serum AAT levels and qualitative analysis by isoelectric focusing, to identify mutations in SERPINA1. Using 3 predictive methods, SVM, FoldX and PolyPhen-2, they categorized the identified mutations as probably deleterious, possibly deleterious, possibly neutral, or probably neutral. The authors also present a benchmarking analysis of SVM predictions against three datasets of known SERPINA1 pathogenic and benign variants from ClinVar.

The study methods are clearly presented and elaborated in the supplementary material. The results are well-written and the key findings of 21 rare/novel mutations identified were presented in detail. The authors provide an important description of how the mutations may result in corresponding AAT levels, especially in the context of combination with more common deficiency alleles. The authors point out a few important limitations such as the fact that this is an observational study and not a controlled study and that they do not report on additional genetic or non-genetic factors that could contribute to the development of the clinical phenotype (COPD).

The conclusions are well supported and suggest that NGS and computational modeling, especially the SVM method, are useful tools to identify important mutations in SERPINA1 that may aid the diagnosis of AATD and lead to improved clinical care for individuals harboring these mutations.

Response: We thank the reviewer for their positive feedback. We have addressed each of the points raised by them and the other reviewers and indicated, when relevant, where changes have been made in the manuscript.

2) Given that AAT inhibits eg. elastase, is there any information on mutations in the elastase gene in these patients that would also influence the progression to the clinical phenotype? For example, mutations in elastase that affect the ability for AAT to inhibit properly?

Response: The reviewer raises an interesting point; however, we are not aware of studies that have looked at the elastase gene specifically, and this is outside the scope of the present study.
3) Is there any additional information regarding presentation of cerebral aneurysm or family history of aneurysm in the patients? ie. were all patients screened for cerebral aneurysms?

Response: There is some evidence in the literature of an association between cerebral aneurism and AAT variants. Thus, this patient was tested for AATD due to this presentation. No further information, i.e., family history of cerebral aneurisms, was available. In response to this comment, we have added some background information on previously reported links between AAT variants and cerebral aneurism (see Discussion, page 17, line 395–397).

4) Are there any cellular or animal models that support that there is an altered functional activity and/or protein levels for the specific SERPINA1 mutations presented in this report?

Response: We are only aware that the I50N variant has been previously tested in a cellular model, and was confirmed as pathogenic – this has been noted in the discussion (see page 17, line 371–373).

Reviewer 3

1) This paper describes the use of computational analyses to understand the growing number of variants of unknown significance (VUS) in genetic disease etiology. With the application of genome and exome sequencing to ascertain the genetic causes behind these diseases this is a problem that is likely to grow. The authors have focused on alpha 1 antitrypsin deficiency (AATD) which is a well recognized disease with multiple clinical phenotypes and this paper uses a cohort that is well described. The analytical methods employed in the paper take into account all aspects known regarding SERPINA1 and don't solely rely on any one metric. This paper is well written and well described and furthers the understanding of the clinical spectrum of genetic variants in AATD. Additionally the critical need for early and accurate diagnosis of AATD is meaningfully pointed out by the authors.

Response: We thank the reviewer for their positive feedback. We have addressed each of the points raised by them and the other reviewers and indicated, when relevant, where changes have been made in the manuscript.
2) The word mutations is used when variants is likely to be what the authors meant (such as p4; line 59 - p5; line 80 - p5; line 84 - p6; line 109 - etc). If these are suspected to be de novo (i.e. mutations) for some reason that should be more clear, otherwise the authors should consider using more general terminology (e.g. variant).

Response: In line with the editor’s comment, ‘mutation’ has been replaced with ‘variant’ throughout the manuscript.