Author’s response to reviews

Title: Genome-wide rare copy number variation screening in ulcerative colitis identifies potential susceptibility loci

Authors:

Hamid Reza Saadati (hsa@informatik.uni-kiel.de; hr.saadati@yahoo.com)

Michael Wittig (m.wittig@mucosa.de)

Andre Franke (a.franke@mucosa.de)

Ingo Helbig (ingo.helbig@uk-sh.de)

Robert Häsler (r.haesler@mucosa.de)

Miles Parkes (miles.parkes@addenbrookes.nhs.uk)

Philip Rosenstiel (p.rosenstiel@mucosa.de)

Stefan Schreiber (s.schreiber@mucosa.de)

Tom Karlsen (t.h.karlsen@medisin.uio.no)

Limas kupčinskas (limas.kupcinskas@lsmuni.lt)

Carl Anderson (carl.anderson@sanger.ac.uk)

Christopher Mathew (c.mathew@umd.ac.uk)

Version: 1 Date: 08 Feb 2016

Author’s response

Author’s response to reviews:

Point-by-point response MGTC-D-15-00134
Reviewer #1

1. I did not understand the reason why you did not study the common CNVs and Ulcerative colitis (UC) considering that there is no published study of common CNVs and UC in the literature.

We completed a systemic genome-wide association analysis of common CNVs and UC previously, but we did not find any confident and replicable association. Therefore, we excluded these results from the manuscript as the analyses raised more questions than providing exhaustive answers. As mentioned in the manuscript, common CNVs (opposed to rare ones) are mainly smaller structural variants, often exist in multi-copy number states and tend to be located in complex repeat-rich and (or) segmental duplicated genomic regions (known as copy number hot spots). These characteristics of common CNVs are the reason for the poor consistency among CNV calls by different CNV prediction and genotyping algorithms and therefore make their ascertainment on array-based platforms less reliable and often biased towards high rates of false positive downstream associations (for more information please see references 9 and 11 in manuscript).

2. It would be great if you can mention clearly the difference in terms of genomic region (kb) between common CNVs and rare CNVs.

Actually there is not a sharp and exact difference in genomic size between common and rare CNVs and this steadily changes as the CNV discovery and genotyping technologies and algorithms continuously improve and as larger population genetic studies are conducted for CNVs. However, as it is mentioned in the introduction and discussion of the manuscript, common CNVs tend to be smaller in size, generally under 25 kb, with a median size of 3 kb and exist in multi-copy number states. In comparison, rare CNVs typically involve larger chromosomal segments (>100 kb) and exist in fewer copy number states i.e. single copy gain/loss (references 8, 9 and 10 in manuscript).
3. I would like to have the reference of this statement in the Introduction on page 5 "However, the disease is mainly triggered in genetically susceptible individuals by environmental risk factors".

It is now added as reference 3.

4. On page 11, Figure 2 is mentioned but I am unable to find that figure.

We have corrected this mistake.

5. Reference 33 don't have any journal or other information.

It is now reference 36 and is provided with the journal and author information.

Reviewer #2

Major changes required:

1. There are several datasets in the study and the sample inclusion scheme is difficult to follow. The authors need to present a flowchart of sample ascertainment, which should clearly show the study populations, previously analysed and newly included samples and inclusion/exclusion criteria. See for instance the flowchart presented in papers: Johnston et al. Am. J. Hum. Genet. 76:609-622, 2005; Hou et al. Nature Genetics 46, 1007-1011, 2014; Park et al. Identification of rare germline copy number variations over-represented in five human cancer types. Mol Cancer. 2015; 14: 25.

We apologize that the sample inclusion scheme was confusing. We have now provided a workflow illustration for the whole study, which includes both the sample inclusion and CNV analysis/filtering scheme.

2. The authors confined their results to only three rare CNVs. It would be worthwhile to present another work-flow chart / pyramid diagram depicting various tires of analyses pipe line and the filtering scheme, i.e., 151 CNVs 24 CNVs 3 CNVs.

Please see our reply to your raised issue No. 1 above.
3. The three identified loci have been mentioned in varying orders in Abstract, Results and Discussion. There should be a consistent sequence for presenting the identified loci.

We changed the order of presenting the three identified loci in the Abstract and we think that they are now mentioned in the whole manuscript in the consistent order as:

Del13q32.1, 15.8 kb single copy loss at chr13: 94,781,525-94,797,285

Dup7p22.1, 119 kb single copy gain at chr7: 5,786,323-5,905,210

Dup8q24.3, 134 kb single copy gain at chr8: 140,390,975-140,524,875

4. There is no mention of the scheme adopted for the selection of UC patients except: page 8, first paragraph…. 'The diagnosis of UC was based on standard clinical, endoscopic, radiological and histological criteria [17].' It would be pertinent to briefly mention the criteria adopted for the characterization of the UC patients, based on scoring method or any other.

We agree with the reviewer that this description is missing and we therefore have added a new subsection under the title of „Patient Recruitment and Ethics” in the Methods section that briefly mentions these criteria.

Minor changes/modifications required:

1. Page 5. Introduction, line 38. Remove the repetition of '……...of the.........'

This has been corrected.

2. Page 6. Introduction, line 52. '….in-silico……..' should be written in italics, as mentioned elsewhere in the manuscript (page 13, line 17).

This has been corrected.

3. There are large number of inconsistencies in the references and many of them do not agree with the BMC Med Genet format.
We have revised the whole reference section and believe that it is now consistent and meet the BMC Med Genet format requirements.