To the editors,

Thank you very much for the opportunity to revise and resubmit our manuscript and thank you to the reviewers for their comments. We have acted on all comments and we believe that we have substantially improved the paper. Our point-by-point responses are provided below.

Many thanks

Madunil Anuk Niriella

Replies to Reviewer's comments

Reviewer ID: 03476575

Manuscript ID: 40964;

The authors attempt to reveal the association of 16 risk SNPs, previously reported in Caucasian populations, with IBD patient in Sri Lanka. A good amount of subjects (200 CD; 214 UC; 465 controls) were recruited for the study from five localities across three major regions. The justification of research intend is valid and could yield meaningful insights onto the predisposing genetic factors of IBD in Sri Lanka. The experimental design and typing method are suitable for a genetic association study.

Despite that, parts of the study are inadequately explained and analyzed, especially for data analysis of the genetic association, which may result in bias interpretation. I also found a number of inconsistencies that requires further formatting throughout the manuscript. I believe that the data in this manuscript is valuable and could fill the gap of knowledge for the mutation prevalence in IBD patients of South Asia, provided that a more comprehensive analysis is to be carried out on the data.

Below are some major comments to the authors:

---

1. Genetic association study should examine both allelic and genotypic distribution for each marker. Authors only analyzed the genotypic frequencies with a *p* value (2 x 3 contingency table on 2 df) as in Table 2. 2 x 2 contingency table with different genetic models (ie, dominant, recessive, additive) should be performed for chi-square test/Fisher's exact test with odds ratios and 95% confidence interval for a more meaning analysis.

We have included this in the revised table where heterozygous and homozygous combinatiosn are analyzed separately and reported as odds ratios and p values.

2. It is unclear that which allele/genotype is risk/protective to the disease.

In table 1, the variant allele is stated as a separate column. This is the allele that is variant from the normal population and causes the risk. The homozygous genotype for this allele has the highest risk and depending on the behavior of the gene the heterozygous may or may not increase the risk.

A separate sentence on selection of SNPs added to the methods section

Are the patients diagnosed according to a set established criteria, such as Vienna classification? If yes, it should be stated in the manuscript and these characteristics should be summarized in Table 1.

This study was not done on newly diagnosed patients. The patients were diagnosed based on clinical and histological criteria by a team headed by consultant gastroenterologist (one of the authors of this study). Hence we are unable to list of classify them based on a specific diagnostic criteria.

However, patients with ulcerative colitis and Crohns disease are further grouped according to Montreal classification. Some relevant categories are listed in table 1. The full phonotypic characteristics (without the genetics component) are published elsewhere (in press).

Are the patients and controls comprised of different ethnicity (Sinhalese, Tamil, Muslim)? If yes, stratification analysis should be performed as genetic heterogeneity is readily known to cause bias in genetic association study.

Included in table 1

Methods [study population]: Authors mentioned that an equal number of unrelated controls were

recruited. But, there are 414 patients vs 465 controls, which is not equal. If it is not 1:1 matching, what is the actual ratio?

The cases and controls were planned to be the same. However when sampling, more controls were recruited. The sentence is and we have removed the 1:1

It would be helpful to include the basic demographic data (eg, age, gender, ethnicity) of the controls in Table 1.

The basic demography of the controls is not available.

Authors could elaborate on the typing methods as it is too brief. The same for result section.

Corrections done to the methods section under typing method used. Result section improed

Table 1: why a significant test is needed (*p* value) since CD and UC are two known subtypes of IBD.

We disagree with this comment. CD and UC are distinct disease entities although categorized under the term "inflammatory bowel disease". These diseases differ with regards to phenotypic characteristics, risk factor and genetic associations. Therefore we believe that the difference between the two groups

Table 2: CD and UC are two known distinct types of IBD, they should be analyzed separately, instead of together.

We have assessed each gene against the controls. Table 3 presents analysis for either CD or UC separately. In table 3 the

Table 2: The total number of cases is 415 here, as opposed to 414 mentioned in the text. Please check. Authors should check the data and numbers in tables carefully. Table 1: the percentage is not tally with the number (eg, family history – 7/200 = 3.5%, instead of 4.65); Duration of disease should be in (mean, SD) instead of (number, %).

Corrections made.

Minor comments: 1. Abstract [methods]: Subject information should be mentioned here.

Corrections made

2. All gene names should be standardized and in italic.

Corrections made

3. Check for consistency throughout the manuscript, eg IL12B/IL12-B; south-asian/south Asian; case-control/case control; p/P/*p*; 100,000/100000; Bonferroni instead of Bon Ferroni.

Corrections made

4. Some keywords may be too lengthy, eg *LAMB1* gene mutation, *IL-12B* gene mutation, genetics of inflammatory bowel disease.

Corrections made

5. There is consistent appearance of unwanted hypen (-) at the end of the rows.

This may have been due to the office software that was used. We hope the current version does not have the same error.

6. Methods [statistical analysis]: "The statistical methods in this study..." this sentence could be moved to acknowledgement, perhaps.

We removed the above sentence. It is described under roles of different authors

7. Table 3: Instead of expressing CD and UC as phenotype, I suggest 'subtype', as phenotype is better fit for clinical characteristics in this case control study.

Changes made

8. Table 3: rs9822268 is located at intron-17 of *APEH* gene, not other gene.

Corrected

9. Table 3: it is unclear that odds ratio in this table refers to which allele/genotype of the marker.

Table 3 revised with odds for hetergenous and homogenous populations seperately, compared to healthy controls

10. Table 4: some characteristics analyzed here are not included in table 1 nor mentioned in the text, eg GIT location of inflammation.

We have now included this