Author's response to reviews

Title: Evaluation of 6 candidate genes on chromosome 11q23 for coeliac disease susceptibility: a case control study.

Authors:

Karen Brophy (brophy4@yahoo.co.uk)
Anthony W Ryan (aryan12@tcd.ie)
Graham Turner (gturner@tcd.ie)
Valerie Trimble (vtrimble@stjames.ie)
Kunal D Patel (kpatel@tcd.ie)
Colm O'Morain (Gastroenterology@amnch.ie)
Nicholas P Kennedy (nick.kennedy@tcd.ie)
Brian Egan (brianeagan76@hotmail.com)
Fiona M Stevens (Fiona.Stevens@nuigalway.ie)
Mohamed Abuzakouk (abuzakm@gmail.com)
Conleth Feighery (Con.Feighery@tcd.ie)
Dermot Kelleher (kellehdp@mail.tcd.ie)
Ross McManus (rmcmanus@tcd.ie)

Version: 2 Date: 23 October 2009

Author's response to reviews: see over
Dear Editor,

Attached is a revised version of the article (MS: 9070282802862730) entitled “Evaluation of 6 candidate genes on chromosome 11q23 for coeliac disease susceptibility: a case control study” by Brophy al., recently reviewed for *BMC Medical Genetics*.

We have performed the revisions suggested by three reviewers, outlined in the pages that follow. Our responses to the reviewer’s comments are indented and presented in italic type, with reference to the location of changes to the revised manuscript.

It is our hope that the article will now be considered suitable for publication in *BMC Medical Genetics*.

Yours sincerely,

____________________________
Anthony Ryan
Reviewer’s report – Reviewer 1

Title: Evaluation of 6 candidate genes on chromosome 11q23 for coeliac disease susceptibility: a case control study

Version: 1 Date: 9 August 2009

Reviewer: Laetitia Michou

Reviewer’s report:

Brophy et al. genotyped tags SNP of 6 candidate genes in cases-controls in celiac disease. Those genes are located in a locus previously reported to be linked to this disease. They found an association with two SNP, supposed to be functional, of IL18 promotor and replicated those results in a second sample.

Major compulsory revisions:

- The “methods” section must be less confusing in several aspects.

Patients and controls: clinical characteristics of patients should be presented (age, sex, age of onset, autoantibodies presence, positive family history of celiac disease, etc…) and demographic characteristics should be presented for controls (age, sex). Same characteristics must be presented for both samples, with comparison between each other to check for homogeneity before pooling. Do both samples belong to the same population?

How do the authors assess the absence of celiac disease phenotype in controls?

We have added an additional table with age and sex data for cases and controls. Furthermore, we have corrected statistically for age and sex using recognized methods implemented in the program HAPLOSTATS (revised table 2, now table 3). We also note that studies of coeliac disease genetics, including the recent Wellcome Trust Case-Control Consortium studies, have not used age and sex matched cases and controls (all UK controls are from the 1958 cohort, while cases are of variable age). Age at onset of symptoms is considered problematic for coeliac disease, as practices in diagnosis have changed considerably over the lifetime of most study subjects, and in many cases the disease is present for many years prior to diagnosis. Our approach has, therefore, been consistent with current practice in the field. In addition, all cases and controls were of Caucasian Irish extraction. The Irish population is relatively homogeneous and has received little inward immigration until very recently.

Genotyping: Do the authors control the quality of genotyping by re-genotyping randomly selected individuals in each sample? The results of Hardy-Weinberg equilibrium check in controls must be presented.

All genotyping included a number (approximately 10%) of anonymous duplicates. A statement has been added to the Methods section. All loci were in HWE in all populations (added to results section)

Statistical analyses: This paragraph must contain information regarding the correction of P-value or not, and regarding the power calculation, that must be performed.

We had stated in the Methods (line 4, line 7 in revised version) and the Discussion (paragraph 2) that uncorrected P values from the first case-control
sample were used to evaluate results for subsequent genotyping. A statement has also been added to the table legend. Furthermore, global haplotype comparisons are significant after correction for multiple testing, and the haplotype comparison has been replicated separately in independent case control sample 2 (Table 3). Text has been added to the results (last paragraph) and the discussion (second paragraph). In addition, a power calculation has been added to the Methods section.

Minor essential revisions:
-Page 5, line 15, do the authors mean “linkage” instead of “association”?

  Yes – this correction has been made.

-For clarification, particularly for Table 1B, controls for the learning sample may be called “Control 1”, those from the replication sample may be called “control 2” and the pool sample of controls may be called “control 1+2”. -Presentation of P-value with only 2 to 3 decimals may improve the reading of tables.

  Changes have been made

Discretionary revisions:
- A comment on the limits of the study could be made in the discussion, regarding the small number of patients (n=150) in the replication sample and regarding the fact that this kind of case-control study does not protect against population stratification bias.

  A brief comment on small sample size has been added to the discussion. Stratification due to ethnic diversity is unlikely to be an issue in an Irish population, which has received very little inward migration until very recently. The Methods section has been modified to note the uniform Irish origin of all cases and controls (Methods, line 3).

- A search for epistasis with the major gene of celiac disease at the HLA-DQ locus may have been performed.

  This study and its conclusions are independent of any association seen on chromosome 6 (HLA). The proposed analysis, while interesting, is not currently possible due to the extent of HLA genotype data for our control samples.

- Search for genotype-phenotype correlations may also have been performed.

  If the reviewer is referring subphenotypes among the coeliac group, the study design has not taken this into account and the data are not amenable to this kind of analysis. Co-occurrence of dermatitis herpetiformis with celiac disease does occur but cannot be excluded for all samples in the current dataset.

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable
**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**
I declare that I have no competing interests
Reviewer’s report – Reviewer 2

Title: Evaluation of 6 candidate genes on chromosome 11q23 for coeliac disease susceptibility: a case control study

Version: 1 Date: 14 August 2009

Reviewer: Abdellatif Maalej

Reviewer’s report:

1- It would be helpful for the reader if the authors could give an estimate of the power of the study. Thus, it would be easier to understand what degree of association could be expected to be detected by the study.

This has been performed using the Genetic Power Calculator

2-In the method subsection, the authors need to give a detailed description of the breakdown of disease (clinical manifestations, immunological data......), the mean age, as well as the male-female ratio of patients and controls.

We have done this within the limitations described above. Age and sex data have been added for cases and controls, and a statistical correction for these has been added to the HAPLOSTATS analysis. Furthermore all samples had biopsy proven coeliac disease, the gold standard for coeliac disease diagnosis.

3-One of the biggest potential pitfalls in case-control studies is the choice of a control group. Where the controls examined? If so how?. Therefore, it is difficult if not impossible for the reader to judge whether an adequate selection of controls was achieved.

Controls were unscreened random population samples with no symptoms of coeliac disease. All control samples were unrelated ethnically Irish individuals.

4-Besides the allelic association test, it is recommended to the authors to use a genotypic association test; in order to compare the genotype distribution in controls and patients the authors can use genotype relative risk test (GRR) (Lathrop GM. Estimating genotype relative risks. Tissue Antigens 1983; 22: 160–162).

Genotype frequencies and P values were included in Table 1 of the original manuscript. The current format allows, therefore, easy calculation of $\chi^2$ and odds ratios. We have commented only on those effects that yielded an uncorrected P-value $>0.05$ in the Results section, where models such as carrier status were included (Results, Paragraph 1). We have further extended this and included an Odds Ratio for the single SNP association.

5- The authors should check for Hardy Weinberg equilibrium for all SNPs.

These tests were performed but not included in Table 1 since, with one exception ($P = 0.04$), all P values were $>5\%$. A statement has been added to the text (Results, paragraph 1)

6-It is not clear why the authors have compared the pool of patients to the second set of controls. I think that they should compare the second set of patients to the second set of controls.
controls (as a test for replication of association) and if the association is replicated then they can compare pooled patients against pooled controls. So I suggest simplifying table 1B by giving only interesting results of tests recommended above.

This has been included for the haplotype analysis (Table 3), comparing case-control 1 and case-control 2 separately.

7- In the statistical analysis the authors should use bonferroni correction for multiple testing.

P values were presented uncorrected. Haplotype analyses using HAPLOSTATS yielded P values <0.001 which are still significant after correction for multiple testing. More detail is given on page 2, in response to reviewer 1’s comments.

Minor comments:
1-It would be better to rewrite and summarise the introduction subsection.

We feel that shortening the Background section will compromise adequate description of all the candidate genes, of which there are 6.

2-Concerning the word “coeliac”, the authors should write it one way “Coeliac” or “celiac”.

This correction has been made (to coeliac)

3-The authors could make the discussion more interesting by introducing other results, as well as reason for differences in the results of different studies.

We have discussed possible reasons for the differences (SNPs analysed or non-detection of a rare haplotype) between our IL18 coeliac results and the results of the other 2 papers to look at this locus in coeliac disease (Discussion, paragraph 5).
Reviewer’s report
Title: Evaluation of 6 candidate genes on chromosome 11q23 for coeliac disease susceptibility: a case control study
Version: 1 Date: 20 August 2009
Reviewer: Holger Kirsten

Reviewer’s report:
The authors described a genetic candidate gene association in celiac disease focusing on six candidate genes within a genomic region known to be linked with celiac disease. Among investigated genetic variants, the authors found evidence for association of 2 SNPs within the gene IL18 with celiac disease. When the authors included a second cohort in the analysis, evidence for association remained for rs1946518, however, the significance of the association decreased compared to the first cohort. Additionally, the authors identified a rare haplotype consisting of these two IL18- SNPs which was significantly associated with celiac disease.

Major Compulsory Revisions
Association analysis of single markers and haplotypes is presented in the manuscript in two different ways. The authors presented data describing single marker analysis for the first cohort and the combined first and second cohort. In contrast, haplotype analysis was presented for all combined cases vs. combined controls and combined cases and controls of the second cohort. As the aim of the validation cohort is to validate the data, haplotype analysis should also be presented for the first cohort separately in addition to the combined cohorts. The value of analysing combined cases with controls of the second cohort remains difficult to understand for this reviewer.

Separate haplotype analyses have been performed for both case-control sets, resulting in the replication, corrected for age and sex, of the haplotypic effect in both datasets.

The authors should report on their initial power analysis. Especially as many markers did not show association with the disease, power analysis will show which effect sizes were detectable within this study.

A power calculation has been added to the methods

In hg19 (UCSC Genome Browser on Human Feb. 2009 Assembly), the markers defining the linkage region on chromosome 11q, D11S4111 and D11S4464 range from chr11:115,35 to 123,13 Mb. Including the location of IL18, the range of interest on chromosome 11 ranges from 112.01 to 123,13 Mb. This genomic range contains (in hg19) 113 genes. As the authors claim, that they performed the first systematic candidate gene investigation on this region, they should give more details, which filter criteria were used to select the six investigated genes out of the 113 genes.

Reference to the 1st systematic candidate gene investigation have been removed

Minor Essential Revisions:
The authors should give more details on how matching, especially ethnic matching, of cases and controls was done.
All cases and controls were ethnically Irish, as outlined above (referee 2).

The authors should state on which genome build genetic coordinates relate to. As stated above, in hg19 (UCSC Genome Browser on Human Feb. 2009 Assembly), markers D11S4111 and D11S4464 define a range on chr11 from 115.35 to 123.13 Mb compared with numbers 115.36 Mb to 123.66 Mb given in the manuscript.

The original analysis was based on papers by Greco et al, which are based on microsatellite maps. Updated locations (GRCh37, Ensembl) have been added to the revised manuscript.

Nomenclature should be more consistent: For example, the authors wrote HLA-DQA1 instead of HLA-DQA1 which additionally should be written in italics (Chapter Background).

HLA-DQ2 was used to refer to the protein, which may be encoded in cis or in trans by 2 separate genes. Gene names are listed using the gene symbol, italicized and presented without hyphens.

There should be a period after et al. (Chapter Background). There should be either “T-cell” or “T cell” throughout the text.

This has been changed

There is another study on MMP3 and celiac disease, which was not mentioned (Hum Immunol. 2005 Jun;66(6):716-20. Epub 2005 Mar 18., PMID 15993717).

This has been included

As the authors of PMID 15993717 described evidence for sex specific association, did the authors found evidence for sex specific association in the 11q data, too?

Since the authors of PMID 15993717 found evidence of male specific association and there is a generally a preponderance of females among celiac disease cases, in the interest of power we have not performed this analysis. However, haplotype statistics are now presented with a correction for age and sex (Table 3).

The authors should give more detail on the observation that “observed TNF levels do not appear to be due solely to genetic variation at the TNF locus (Daly et al., unpublished).” Who are et al. in this context?

A citation has been added

The authors claim, “To date, no genetic association studies have been carried out to test for association between polymorphisms in the IL10RA gene and celiac disease.” This is not correct, as in 4. van Heel DA, Franke L, Hunt KA, Gwilliam R, Zhernakova A, Inouye M, Wapenaar MC, Barnardo MC, Bethel G, Holmes GK et al: A genome-wide association study
for celiac disease identifies risk variants in the region harboring IL2 and IL21. Nat Genet 2007, 39(7):827-829. Illumina HumanHap300 Bead chips were used. This includes probes for rs4252287 rs2229113 and rs9610 which are situated within L10RA.

This has been changed in the text.

In general, the authors should give a direct comparison how the analysed six genes performed in the genome wide association study.

Addition to text, Discussion "Four of the 16 SNPs used in this study have been included in a genome-wide analysis of coeliac disease risk [4], where none showed association with disease. While that study did include analysis of IL18-607 (rs1946518), IL18-137 (rs187238) was not analysed. The haplotypic found in this study would not, therefore, have been detected”

What are the parameter used for the tagging procedure? With which efficiency are which SNPs tagged?

SNP data from Perlegen, dbSNP or where possible sequence data from SeattleSNPs or InnateImmunity were used to tag common haplotypes (>5%) for SNPs with MAF>0.05. This has been added to the methods section (2nd paragraph).

The authors stated that “[SNPs] with a theoretical or proven functional role (e.g. gene promoter regulation or transcription factor binding site alteration) have been included” These SNPs should be labelled in Table 1.

This is the case with the IL18 promotor SNPs which, as stated, appear to influence expression levels. This is discussed in the Background section. An asterisk has been added to Table 2 (Table 1 in original submission) to indicate the putative functional nature of these variants.

The authors should comment on the fact, that evidence for single marker association of both IL18-SNPs decreased when cohort 2 was included in the analysis.

This has been added to the text, Results, second paragraph (“though weaker than that observed for celiac 1 Vs control1”)

Which functional test could be done in the future in order to understand the role of the IL18 disease-haplotype in celiac disease?

While we agree that functional follow-up is interesting, we feel it is beyond the scope of the present study

A graphic showing the LD pattern of IL18 would be very valuable.

This can now easily be constructed from publicly available datasets and would occupy considerable space if added to the current manuscript, without
significantly adding to the current paper. We have noted in the Discussion (Paragraph 4) that the LD patterns are similar to those observed for the Hapmap CECPH Europeans.

Discretionary Revisions
A supplementary graphic showing the LD pattern of all genes would be valuable.

Reference to LD across the 11q region is mentioned in the Discussion, paragraph 4.

Level of interest: An article whose findings are important to those with closely related research interests
Quality of written English: Acceptable
Statistical review: No, the manuscript does not need to be seen by a statistician.
Declaration of competing interests:
I declare that I have no competing interests