Author's response to reviews

Title: Characterization of the association between 8q24 and colon cancer: gene-environment exploration and meta-analysis

Authors:

Carolyn M. Hutter (chutter@fhcrc.org)
Martha L. Slattery (marty.slattery@hsc.utah.edu)
David J. Duggan (dduggan@tgen.org)
Jill Muehling (jmuehling@tgen.org)
Karen Curtin (karen.curtin@hsc.utah.edu)
Li Hsu (ljih@fhcrc.org)
Shirley A. A. Beresford (beresfrd@u.washington.edu)
Aleksandar Rajkovic (rajkovic@upmc.edu)
Gloria E. Sarto (gsarto@cwhr.wisc.edu)
James R. Marshall (james.marshall@roswellpark.org)
Nazik Hammad (hammadn@karmanos.org)
Robert Wallace (robert-wallace@uiowa.edu)
Karen W. Makar (kmakar@fhcrc.org)
Ross L. Prentice (rprentic@whi.org)
Bette J. Caan (bette.caan@kp.org)
John D. Potter (jpotter@fhcrc.org)
Ulrike Peters (upeters@fhcrc.org)

Version: 3 Date: 18 September 2010

Author's response to reviews: see over
September 17, 2010

Dear Dr. Bruecher:

Please find enclosed the second revision of our manuscript MS: 1751336102334254 “Characterization of the association between 8q24 and colon cancer: gene-environment exploration and meta-analysis” by Hutter et al. which we are resubmitting in response to reviewers for publication in BMC cancer.

We appreciate the feedback and reviews we received. We are glad to see that the Referee 1 reports “The paper is acceptable now” and that Referee 2 had no additional comments and called our paper “An article of outstanding merit and interest in its field”. We thank the third referee for his review of our paper. In response to that review we have made several modifications to our manuscript. We used track-changes in Word to highlight the changes, and present a point-by-point response below. Note the page and line numbers provided refer to the uploaded version with track-changes in the text.

Thank you for considering our manuscript for publication in your journal. We look forward to hearing your decision on this paper.

Sincerely,

Carolyn M. Hutter, PhD
Staff Scientist
Fred Hutchinson Cancer Research Center

**Referee 3**

**Reviewer's report**

**Title:** Characterization of the association between 8q24 and colon cancer: gene-environment exploration and meta-analysis  
**Version:** 2 **Date:** 12 July 2010  
**Reviews:** Elias Zintzaras

**Reviewer's comment 1:**  
The WHI and DALS studies cannot be combined in the context of a general linear model due to different design settings. Thus, only separate analysis for each study should be shown.

**Response 1:**  
We agree with the reviewer that, since the studies have different design settings, one cannot simply pool the results for WHI and DALS into one generalized linear model. However, we wish to clarify that is not the approach that we used. We used a two-staged approach that first performs the appropriate analysis within each study independently, and second uses a random effects meta-analysis to combine the results. This method is well established, and has been shown to be valid for combining studies with different designs [1,2]. We apologize for any confusion due to wording in our original version, and hope our revision makes this clear.
We also recognize that the reviewer, and other readers, may be interested in seeing the results for WHI and DALS separately, as well as the combined results. We now present the results for each study as part of table 2 (rather than having that as an additional file).

In response to this comment we have made the following changes:

a. We have changed wording in abstract, methods and results to clarify our two stage approach:
   Abstract Page 3 lines 6-7:
   Analysis was performed separately within each study, and combined using random effects meta-analysis.
   Methods, Page 5 line 12:
   The study uses colon cancer cases and controls from two previously described study populations:
   Page 9 lines 14-18:
   To estimate the association between genetic variants in the 8q24 region and colon cancer risk we calculated odds ratios (ORs) and 95% confidence intervals (95% CIs) using a two-stage pooled analysis, which allow us to combine results for the two studies, even though they have different designs [42,43]. Stage 1 used a model appropriate to each study design with study-specific confounders. Stage 2 used a random effects meta-analysis to combine the results for the two studies.

b. We provide the separate analysis for each study as part of table 2 in the main document (previously we only had this information in the supplemental materials).

Reviewer’s comment 2:
The genotype distribution of each variant should be presented and an analysis based on contrasts of genotypes (dominant, recessive, additive, co-dominant) should be performed. The analysis based on the allele contrasts always increases the chance of significant results. An additional analysis based on the generalized OR may reveal hidden associations, especially for the tumor stage analysis (see Zintzaras E. The generalized odds ratio as a measure of genetic risk effect in the analysis and meta-analysis of association studies. Stat Appl Genet Mol Biol. 2010;9(1):Article21.).

Response 2:
Our primary analysis was based not on allele contrasts, but on log-additive model with genotypes coded 0/1/2. However, we agree with the reviewer’s point that the genotype distribution of each variant should be presented, and we have modified table 2 accordingly. Specifically, we give the genotype distribution of each variant in each study, and present the unrestricted contrasts (comparing the heterozygotes to the reference homozygotes, and the non-reference homozygotes to the reference homozygotes), as well as the log-additive model OR and p-value. We also examined dominant and recessive models for each variant, and mention the results in the text.

We examined the tumor stage analysis using the reviewers generalized odds ratio method [3]. Similar to our polytomous regression results, none of the analysis were significant when we used the generalized OR to examine stage.
In response to this comment we have made the following changes:

a. We have modified table 2 to give the genotype distribution and the unrestricted (as well as log-additive) genotype contrast results.

b. We examined dominant and recessive genotype contrasts, and describe this in the text:
   Methods, page 9 line 25- page 10 line 1
   We also examined each SNP using a dominant and recessive model
   Results, page 13 lines 9-12:
   We did not observe any additional significant SNPs when we considered dominant or recessive modes of inheritance

c. We used the generalized odds ratio [3] for the tumor stage analysis, and describe this in the text:
   Methods page 10 lines 21-22
   In addition to polyomous regression, we also used the generalized odds ratio to examine stage, ranking localized, regional and distant disease in that order [46].
   Results, page 14 line 9-10:
   Our stage analysis was also not significant when we used the generalized OR.

Reviewer’s comment 3:
In the WHI and DALS studies, there are 126 and 139 non-whites, respectively, but the populations were in HWE, an explanation is needed for this result.

Response 3:
Our original statement that “allele frequencies among controls did not deviate from HWE” applied only to the subjects who self-reported as white. We apologize for not explicitly stating that. The reviewer does raise the point that readers might be interested in seeing the HWE values for both the whole population, and for subjects who self-report as white. We now provide that as part of our additional files. We do note, that even when we include the non-whites, we still do not see large deviation from HWE. Further, the four SNPS of primary interest (rs10505477, rs10808555, rs6983267 and rs10956368) are all in HWE in both populations, even when we include the non-whites. This is because HWE is not a powerful test, and it has been shown that the power for detecting admixture using HWE is generally small [4].

In response to this comment we have made the following changes:

1. We have added a supplemental table that gives the HWE p-values for controls for both the whole population and for subjects who self-report as “white”. These p-values are presented separately for WHI and DALS.
2. We have clarified that our examination of HWE was for subjects who self reported as white:
   Methods Page 8 lines 10-12
   the allele frequencies among controls for self-reported whites did not deviate from HWE for any SNP at p = 0.01. HWE for all SNPs for both the self-reported whites and the full population are presented in Additional file 1.
   Results Page 12 line 21-22
   We genotyped 11 SNPs, and all were in HWE within the control population of subjects who self-reported as white in both studies.
**Reviewers comment 4:**
The significant results of the subgroup analysis might be data driven since the study protocol was not designed to investigate these specific subgroups. Furthermore, the study was not designed to investigate gene-environment interaction. All risk factors can be used to adjust the genetic risk effect.
The analysis and presentation of the mass subgroup results corresponding to risk factors raises the need for multiple testing adjustment.

**Response 4:**
We agree that any findings in the subgroup analysis need to be interpreted in light of the fact that we examined 11 environmental factors, stage and site.

Although we did not do a formal multiple testing adjustment, we do not feel that we have overstated these findings. In our abstract we had stated:

“We did not observe any notable evidence of effect modification by known colon cancer risk factors, and risk did not differ significantly by tumor site or stage”

And we did not call the subgroup findings with p<0.05 “significant. We did note that SNPs showed slightly stronger evidence for association in men vs women, but noted “this is not a statistically significant difference in either DALS alone, or in the combined sample”. When discussing the differences in stage and site, we use terms “slight tendency” and “potential association”, but noted that “these differences were not larger than would be expected by chance”.

However, we acknowledge that reviewer raises a valid point, and have made several wording changes to stress even more strongly, that our findings need to be interpreted in the context of the number of factors examined, and the exploratory nature of such an examination of gene-environment interactions. We stress, however, that we do not think that this changes the primary message or interpretation of this paper.

That said, we do want to state that we think the study populations used are appropriate for this type of analysis. Both studies are well-characterized and collected detailed information on epidemiologic risk factors, as well as blood to be used in genetic analysis. Just because it was not a primary initial goal of the parent studies, this does not invalidate examination for gene-environment interactions using these study populations.

In response to this comment we have made the following changes:

1. We have revised our wording to make it clear that our results need to be interpreted in the context of the number of SNPs and interaction factors examined.
   Abstract: we changed wording from 
   *We observed statistically significant associations* to
   *we observed evidence for associations* 

   We also reworded the sentence starting with “within this region…” to now read
   *The combined results for our two studies of colon cancer showed an OR=1.10 (95% CI: 1.01-1.20, \( P_{\text{trend}}=0.023 \)), and a meta-analysis of our results with previously reported studies of colon and colorectal cancer strongly support the association for this SNP (combined OR for rs6983267=1.21, 95% CI: 1.18-1.24, \( p=5.5\times10^{-44} \)).*
However it is important to evaluate our findings in light of the number of SNPs tested. A conservative Bonferroni correction would for 11 SNPs would lead to a p-value threshold of 0.05/11=0.0045.

None of the SNPs would be significant if we used a correction for multiple testing for the 11 SNPs. But this finding is not significant in either subsite once you adjust for multiple testing (see Additional file 5).

Although some trends were observed, we acknowledge that we have examined a large number of factors, and our results should be interpreted in the context of the large number of tests performed.

**Reviewer’s comment 5:**
Haplotypes frequencies should be estimated and the haplotypes of cases and controls should be compared.

**Response 5:**
At the reviewers suggestion, we performed haplotype analysis and found no significant associations.

In response to this comment we have made the following changes:
1. We describe our use of haplotype tests in the methods section and present the results in the text:

**Reviewer’s comment 6:**
The significance of the association results should be adjusted be the number of examined variants.

**Response 6:**
We addressed this comment in our response to point 4 above.

**Reviewer’s comment 7:**
A high LD in HapMap does not justify the switchability between two variants.
Response 7:
We clarify that we chose to present results for both rs6983267 and rs10505477 because they are in high LD and some published studies have used rs10505477 as a tag for rs6983267. However, our primary result (in the abstract, first paragraph of the discussion, etc) is for the rs6983267 alone. In order to not detract from this primary finding, we have moved the figure with results for the combined analysis of both rs6983267 and rs10505477 to the additional files/supplemental material.

In response to this comment we have made the following changes:
1. We moved the forest plot from the analysis of both rs6983267 and rs10505477 from the main text of the paper to the supplemental materials.
2. We reworded our justification for also including rs10505477

Methods page 10 line 25-page 11 line 4
Because rs6983267 and rs10505477 are in high linkage disequilibrium (LD; r²=0.93 in the HapMap CEU population), and because some studies have used them as tags for one another [32], we also searched for studies that examined the rs10505477 variant. Since it may be argued that high LD in HapMap does not justify switchability of the two alleles, we present the results for rs6983267 only as our primary analysis.

Reviewer’s comment 8
In the meta-analysis use the RE model.

Response 8:
We have taken the reviewers advice and now use a random effects model for the meta-analysis.

In response to this comment we have made the following changes:
1. Methods page 11 lines 12-14
   Risk estimates from individual studies were combined and the corresponding summary 95% CI and p-values were obtained under a random-effects meta-analysis model [47,48]
2. All results and figures for meta-analysis were updated to reflect use of random effects meta-analysis

Reviewer’s comment 9:
The Egger’s test and the funnel plot do not really investigate publication bias (see Zintzaras and Lau 2008 J Clin Epidemiol).

Response 9:
We have clarified our discussion to emphasize that the Egger’s test and funnel plot compare results of large and small studies, rather than explicitly testing publication bias. We also reviewed the reviewer’s paper (as suggested) and have added a cumulative meta-analysis as part of our evaluation of potential bias and heterogeneity.

In response to this comment we have made the following changes:
1. We now include a cumulative meta-analysis plot in addition to the forest plot and funnel plot.
2. We site the Zintzaras and Lau 2008 J Clin Epidemiology paper.
3. We modified the wording used in the presentation of the funnel plot and Egger’s test, and added text on cumulative meta-analysis:

Methods page 11 lines 19-21
Potential bias was assessed by comparing results for small and large studies using Egger’s test and visual inspection of funnel plots [51], and we examined the trend in risk estimates over time using cumulative meta-analysis [52]

Results: page 14 lines 4-6
There was no evidence for differences between large and small studies based on Eggar’s test (p=0.44) and visual inspection of funnel plots. The cumulative meta-analysis shows that while the original estimate is slightly higher, the estimates have held steady in the range of 1.21 to 1.23 (see Additional file 7).

Discussion page 16 lines 12-16
There is no strong evidence for reporting bias for this meta-analysis based on both funnel plots and cumulative meta-analysis, however it is still possible that publication bias is preventing, or delaying, the publication of studies that are not replicating the association[52]. It is also possible that there is substantive bias, even in the presence of a symmetrical funnel plot [53].

Reviewer’s comment 10:
A meta-analysis that includes the WHI and DALS studies plus other published studies should be performed for all 11 variants (whenever published studies with these variants exist).

Response 10:
We agree with the reviewer that this would be an interesting and informative analysis. However, at the current time none of the other variants have been examined in >3 studies. With so little evidence per study a meta-analysis is premature and uninformative. We do think this would be important in future examination of this region, and make mention of this suggestion in the paper.

In response to this comment we have made the following changes:
1. We address this concept in the discussion.

Discussion page 17 lines 19-22:
In this paper we focused our meta-analysis on rs6983267, because it is the most frequently published SNP, and the only variant with more than three publications. If studies of 8q24 SNPs other than rs6983267 continue to be published for colorectal cancer, it will be informative to perform meta-analysis of these other variants.

Reviewer’s comment 11:
In Table 1, P-values should be provided.

Response 11:
We respectfully disagree with the reviewer on this suggestion. Our table 1 is a descriptive table with the purpose of allowing readers to examine features of cases and controls across the two studies. We refer to the STROBE guidelines to point out that it is not appropriate to use significance testing or to present p-values in this type of table.[5] Specifically the STROBE guidelines state:
“Inferential measures such as standard errors and confidence intervals should not be used to describe the variability of characteristics, and significance tests should be avoided in descriptive tables.”- STROBE guidelines Vandenbroucke, et al. Strengthening the Reporting of
References used in response to reviewer:


Deng HW, Chen WM, Recker RR: Population admixture: Detection by Hardy-Weinberg test and its quantitative effects on linkage-disequilibrium methods for localizing genes underlying complex traits

