Author’s response to reviews

Title: IL6 and CRP haplotypes are associated with COPD risk and systemic inflammation: a case-control study

Authors:

- Dilyara G Yanbaeva (d.yanbaeva@pul.unimaas.nl)
- Mieke A Dentener (mieke.dentener@pul.unimaas.nl)
- Martijn A Spruit (martiinspuit@proteion.nl)
- Jeanine J Houwing-Duistermaat (j.j.houwing@lumc.nl)
- Daniel Kotz (d.kotz@hag.unimaas.nl)
- Valeria Lima Passos (valeria.limapassos@stat.unimaas.nl)
- Emiel FM Wouters (e.wouters@lung.azm.nl)

Version: 2 Date: 12 December 2008

Author’s response to reviews: see over
Dear Editorial Team,

Thank you for your letter from November 13th regarding our manuscript entitled “IL6 and CRP haplotypes are associated with COPD risk and systemic inflammation: a case-control study”. We kindly accept the opportunity to resubmit a revised version of the manuscript.

We thank the reviewers for their positive and constructive comments. After careful consideration of all recommendations, we have adjusted the text and tables to make it clear for readers. All changes are highlighted in yellow to facilitate your review. In addition, point-by-point answer to the comments is provided.

We hope that the paper in its present form will be accepted for publication in BMC Medical Genetics.

Yours sincerely,
On behalf of all authors,
Dilyara Yanbaeva
POINT-BY-POINT RESPONSE TO REVIEWS

Reviewer: Jemma Wilk

Reviewer's report:

The authors have evaluated association between SNPs and haplotypes in 3 genes (CRP, IL6, and FGB) with COPD, the corresponding biomarkers, and COPD-related phenotypes. They have produced a nicely written paper with interesting findings.

Authors’ reply: The authors thank the referee for her positive reaction and compliments.

Major Compulsory Revisions:

1) There is some discordance between the statements made in the methods/results and the presentation in the tables. The authors indicate that COPD and the 3 measured systemic inflammation levels were treated as 4 primary outcomes. However, in both Table 5 and Table S5, there are results labeled as “Not tested”. The authors appear to have tested IL6 SNPs and haplotypes against all 4 outcomes, but CRP and FGB is only tested against COPD and the corresponding biomarker. The rationale for this is unclear.

Authors’ reply:

Thank you for drawing our attention to this point. In the original manuscript an explanation why we have not tested all possible associations was placed to the Supplement (section Statistical analysis). A rationale for genetic association tests was based on known biological pathways of CRP, IL-6 and fibrinogen. First, we aimed to test each gene against its direct product (corresponding plasma biomarker) and COPD. Second, given that IL-6 is a major and strong regulator of hepatic production of CRP and fibrinogen, we also tested associations of IL6 haplotypes with CRP and fibrinogen plasma levels. Four other associations (CRP haplotypes and IL-6 and fibrinogen levels, FGB gene and CRP and IL-6 plasma levels) were not tested because of two reasons: 1) a little literature evidence of their possible relationship 2) to reduce spurious association by limiting number of tests. To clarify this, the following text was placed to the Main paper Statistical analysis, section Genetic association analysis: Eight primary genetic tests were performed. First, each gene was tested for association with two major outcomes: either corresponding protein level or COPD. These comprised 6 tests. Additionally, given that IL-6 is a major regulator of hepatic production of CRP and fibrinogen, we also investigated the association of IL6 haplotypes with CRP and fibrinogen plasma levels (2 tests)

2) On page 10, the authors indicate they “calculated haplotype-specific scores and then estimated haplotype effects…”, but it would appear that they only performed haplotype-specific analysis when the global haplotype test achieved a p-value < 0.1, which is quite reasonable. If this is correct, then that should be stated. If this is the
rationale, then it would appear that the CRP haplotype-specific analyses should be performed for the phenotype Maximum workload, watts (Table 8).

**Authors’ reply:**

Thank you for these comments. Indeed, we have performed haplotype-specific analyses if global haplotype score test was below 0.1. We have added this statement to the Statistical analysis section in the Main paper (subsection Genetic association analysis).

Regarding further haplotype-specific analysis of the CRP gene for the phenotype maximum workload, we have already performed the haplotype-specific analysis when preparing the original manuscript, but found only marginal association of the haplotype 121211 with maximum workload (P=0.05). To make it clear in the paper, we have modified the Results section as follows: *For all 3 genes, no significant association was observed when overall test of association was applied, except for a marginal association of the CRP gene and maximum workload (Table 7). Further testing has shown borderline association of the haplotype 121211 with maximum workload (P=0.05).*

3) The authors should discuss the limitations of their FGB analysis and results given the failure of 2 tag SNPs for genotyping.

**Authors’ reply:**

Two out of 6 FGB SNPs failed for TaqMan assay development because they are located in repeat rich regions: rs2227432 in MIR region in the FGB intron, rs2227439 in AluSg region of 3’ flank area of the gene. As a result of the failure of these tag SNPs for genotyping, we were not able to discriminate between wild type haplotype 111111 and a common haplotype 111121 defined by rs2227439. The latter haplotype was shown to have 10% frequency in the Dutch population (Uitte de Willige et al, 2005, reference 19 in the main paper). Regarding the tag SNP rs2227432, it was shown to be a part of a rare haplotype 112112 (2% in Dutch controls in the study of Uitte de Willige, 2005) and most likely would not influence common haplotypes estimation. The following sentence was added to the Discussion: *As a result of the failure of 2 FGB tag SNPs for genotyping we were not able to discriminate one common (defined by rs2227439) and one rare (carrying the minor allele of rs2227432) haplotypes of the FGB gene (Uitte de Willige, 2005). They were non-discriminative from the FGB haplotypes 1111 and 1121 respectively.*

**Minor Essential Revisions:**
1) Please add a comment about the genomic locations of the 3 genes you are studying.

**Authors’ reply:**
We have added information on gene location to the Table 3.

2) In table 1, please provide specific p-values (such as with scientific notation) as opposed to 0.000.
Authors’ reply: Correction is provided as follows: “<0.001” instead of “0.000”

Discretionary Revisions
1) The rs2069849 SNP in IL6 that was excluded due to low minor allele frequency was selected as a tag SNP and appears to have similar, if not increased, frequency as compared to HapMap CEU. It is a synonymous SNP. It seems worthwhile to study.

Authors’ reply:
Indeed, in our study the synonymous SNP rs2069849 in IL6 gene has a slightly higher frequency (2% in controls) than in HapMap CEU (1%). It was selected for genotyping as a tag SNP because its frequency of 7% in our reference panel, the PGA-EUROPEAN-PANEL (Seattle SNPs project). However, after genotyping was performed, it turned out that allele frequency of the rs2069849 is below threshold selected in the study (5%), which is based upon study sample size. There was also no significant difference in allele and genotype frequencies between cases and controls. Based on these facts, we have decided to exclude the SNP from the haplotype analysis. Even if the SNP rs2069849 would be included in the haplotype analysis, the haplotype tagged by a minor allele of the SNP, will not achieve the threshold for common haplotypes (frequency>2.5 %). To put some extra information on rs2069849, we have added results of genotype-based analysis to the Table 3 in the Main paper (next to P-value from the allele test).

Reviewer: Craig Hersh

Reviewer's report:
General comments:

The authors have performed a modest-sized case-control genetic association study, examining three genes for inflammatory markers, as well as the corresponding protein levels, as they associate to COPD risk and COPD phenotypes. They find an association between an IL6 haplotype and COPD risk. Interestingly, this haplotype is associated with lower CRP levels. Multiple other analyses are presented in the paper; this apparent lack of focus seems to detract from this association and some of the other interesting results.

Authors’ reply:
We greatly appreciate the interest of the referee in our work and thank him for reviewing the manuscript. The manuscript was adjusted according reviewers’ comments. The revised version of our paper contains less tables and a clear description of the statistical analyses.

Major compulsory revisions:

1. The most striking result from the biomarker analysis is that 40% of COPD patients did not have evidence of systemic inflammation. However, few additional details on this potentially important COPD subgroup are provided. For example, a comparison
of clinical characteristics (as in Table 1) between the “inflammatory” and “non-inflammatory” COPD patients should be included.

Authors’ reply:

In order to select “inflammatory” and “non-inflammatory” subgroups, we have dichotomized COPD patients based on CRP levels (<3 and >3 mg/L) and added an additional table with basic clinical characteristics of the “inflammatory” and “non-inflammatory” groups (Table 2 in the revised Manuscript). It turned out that patients with baseline CRP levels more than 3 mg/L are older, more likely to be male, have higher BMI, more co morbidities and walk fewer meters in 6 minutes. Although we do agree that presentation of the results in two groups (“inflammatory” versus “non-inflammatory”) are illustrative, those clinical parameters were shown as independent significant predictors of CRP levels in a linear regression analysis (former Table 3, which has now been moved to the supplement (Table S2)).

The cut-off point 3 mg/L was selected based on recent publications of de Torres et al (Chest 2008, 133:1336-1343) and Dahl et al (Am J Respir Crit Care Med 2007, 175:250-255). We consider that the selection of a cut-off point which was already used in several COPD studies will help to perform results evaluation across different studies. We have put additional text to the Results to describe the new table.

2. If the non-inflammatory patients seem to represent a distinct COPD subgroup, based on clinical characteristics, then a genetic comparison between the “inflammatory” and “non-inflammatory” COPD patients would be justified.

Authors’ reply:

We understand that the referee wants a genetic comparison between “inflammatory” and “non-inflammatory” groups. However, we want to be precautionary since comparison of clinical characteristics between “inflammatory” and “non-inflammatory” COPD subgroups does not support the conclusion that those groups represent distinct clinical subgroups. First, the “inflammatory” group is older and consists of more males. It is well known that CRP levels are higher in males (in our study median CRP in male patients is 5.4 and in female patients is 2.9) and increase with age (Kathiresan et al, Circulation 2006, 113(11):1415-23). Second, patients with higher CRP had higher BMI and more self-reported co morbidities. These could affect their physical performance such as in 6 minute walking test. Furthermore, due to the cross-sectional nature of our study, we cannot make causal conclusions regarding relationship between clinical parameters and inflammatory markers levels. There are also several statistical issues which prove preferable the utilization of quantitative variable rather than dichotomized. The potential negative consequences of dichotomization are lack of statistical power, decreased validity of the analysis, and loss of generalizability of results (Ravichandran and Fitzmaurice, Nutrition 2008, 24(6):610-1; MacCallum et al, Psychological Methods March 2002, 7(1):19-40). These drawbacks can far outweigh any resulting simplification of the analysis. On these grounds, we have selected a full scale of markers’ variability, assessed by high sensitive assays. We consider this alternative gives a more refined characterization of the data as compared to the selection of binary inflammatory outcome.
3. In general, the SNP selection strategy and statistical analyses are appropriate. However, despite the authors' claim to the contrary, spurious association due to multiple testing is a concern, given the testing of 3 genes, 3 biomarkers, COPD risk, and 6 other COPD phenotypes (Table 8). Given this, the rationale for the Bonferroni p-value of 0.05/8 is not clearly explained. The haplotype-based analysis offers some protection against multiple testing, but replication in a second cohort would be ideal, especially for the association between IL6 and COPD risk. However, interesting results from the two revisions above might outweigh the limitation of the lack of a replication cohort.

Authors’ reply:

First of all, we are grateful for the recognition of statistical analysis standards of our study. Bonferroni P-value of 0.05/8 was selected based on 8 primary tests for major outcomes protein levels and COPD (every gene was tested for its own protein and COPD, 6 tests, and the IL6 gene in relation to CRP & fibrinogen levels (2 tests)). To clarify this, the following explanation was added to the Methods the Main paper Statistical analysis, section Genetic association analysis: Eight primary genetic tests were performed. First, each gene was tested for association with two major outcomes: either corresponding protein level or COPD. These comprised 6 tests. Additionally, given that IL-6 is a major regulator of hepatic production of CRP and fibrinogen, we also investigated the association of IL6 haplotypes with CRP and fibrinogen plasma levels (2 tests). To reduce the problem of spurious associations due to multiple testing, Bonferroni correction was applied for those tests (P_corrected=0.05/8 tests=0.006).

Six other COPD phenotypes were tested as secondary outcomes after initial analysis was performed and were not considered for Bonferroni correction calculation. From other study (Wijkstra et al, Thorax 1994, 49(5):468-72, Cote et al, Chest 2007; 132:1778-1785) and from the current study it is known that COPD functional phenotypes are highly correlated with each other. For example, strong correlations were found between FEV1 and workload (r=0.66), 6MWD and workload (r=0.62), MRC score and 6MWD (r=-0.46) and FEV1 and 6MWD (r=0.4). In this case Bonferroni correction would be too conservative to apply, because it assumes independent outcomes testing.

To address this in the paper, the text was modified as follows: "Body-Mass Index, Airflow Obstruction, Dyspnea, and Exercise Capacity (BODE) index, 6 minute walking distance (6MWD), maximum workload at the cardiopulmonary test, Medical research council (MRC) score, body mass index (BMI) and forced expiratory volume in 1 sec (FEV1) were tested after initial 8 analyses have been performed and treated as secondary outcomes. Secondary COPD outcomes were not considered for a conservative Bonferroni correction, because of a high degree of correlation with each other."

We agree with the referee that replication in the second cohort is essential for a genetic study. However, our manuscript is a first to report a study of haplotypes
tagging genetic analysis covering common gene variation in CRP, IL6 and FGB in relation to COPD. Further studies are required to replicate our data.

4. Partly because of the multiple analyses, the paper seems a bit long, with too many tables. For example, Table 8 could be eliminated, as all results are non-significant. And Table 4 might be combined with Table S4, which both report the SNP-based analyses. If the focus is truly on the genetic results, then Tables 2 and 3 also seem excessive.

Authors’ reply:

Based on the referee comments, we have reorganized tables to make the paper more convenient for readers. Former Tables 2 (Correlation between inflammatory markers and functional measures of disease) and 3 (Clinical predictors of plasma CRP, IL-6 and fibrinogen levels in COPD patients which retained in the best-fit multiple linear regression models) were placed to the Supplement (Tables S1 and S2 in the revised paper). SNPs’ characteristics, results of allele based (former Table 4) and SNP based analysis (former Table S4) were combined in Table 3. Table 7 (former Table 8) could not be eliminated because there is a one borderline association (P=0.06 for CRP haplotypes and workload) which was tested further, although produced non-significant results. Please, see our response to Jemma Wilk for detailed explanation (major compulsory revision #2).

5. A recent paper reports association between IL6 SNPs and COPD risk (Cordoba-Lanus, Respiratory Medicine 2008). Please compare the current results to this previous paper. This may help with the issue of replication in comment 3.

Authors’ reply:

Thank you for drawing our attention to the paper of Cordoba-Lanus et al published in Respiratory Medicine in December 2008. They have studied three SNPs in the IL6 gene: rs1800795, rs1800797 and rs1800796. The haplotype tagged by the latter SNP had a lower prevalence in COPD patients of Spanish origin as compared with the healthy controls and current smokers. Unfortunately, the SNP rs1800796 was not selected for current study due to the fact that it was not genotyped in the reference panel - Seattle SNP project. In HapMap CEU reference panel it has a frequency of 4% (still below our study threshold of 5%). Regarding first two SNPs, rs1800795 and rs1800797, public databases show that they are almost in complete linkage disequilibrium (LD) in European population. Similar LD (D’=0.96) was shown in the study of Cordoba-Lanus et al. We have also studied the rs1800797 and, like our Spanish colleagues, could not find any association between this SNP and COPD susceptibility. Two previous papers (references 44-45 in the Main paper) could not detect any association between the SNP rs1800795 and COPD. Therefore, results from 4 studies are in line for a big LD block consisting of 8 SNPs (including rs1800795 and rs1800797). To address this issue the following text was included into the Discussion: Similar to our results, three previous studies of the IL6 gene could not find any association between COPD and the SNPs rs1800795 and rs1800797 [43-45]. Interestingly, a new association was found with the IL6 SNP rs1800796 in COPD
patients from Spain [45]. Because of a low allele frequency in Europeans (4%), this SNP was not selected for current study; however, further studies using large samples size of Dutch COPD patients are required to replicate this new association.

Minor essential revisions:

Table 1: I assume p-values listed as “0.000” mean p<0.001. Please correct.

Authors’ reply: Correction is provided as follows: “<0.001” instead of “0.000”

Discretionary revisions:

1. Table 6: Including the p-values from score test and from logistic regression is confusing. Might be better to present one or the other.

Authors’ reply:
Based on the referee suggestion, we have reorganized Tables 5 and 6. We have chosen for presentation of the results from logistic regression analysis only.

2. Tables S1-S3 might be replaced by corresponding LD figures. This would help to limit the number of tables.

Authors’ reply:
We have combined Tables S1-S3 into one (Table S3). Unfortunately, software producing LD figures does not accept tri-allelic SNP information to make nice pictures of LD structure of CRP gene. Because of that, suggestion of the referee cannot be followed.