Reviewer's report

**Title:** Associations of ECP (eosinophil cationic protein)-gene polymorphisms to allergy, asthma, smoke habits and lung function in two Estonian and Swedish sub cohorts of the ECRHS II study

**Version:** 2  **Date:** 11 March 2010

**Reviewer:** Jon Genuneit

**Reviewer's report:**

In this manuscript the authors describe the association between three polymorphisms in the ECP gene and allergy, asthma, smoking, and lung function. In particular, the authors try to give indication of effects in subgroups of disease or smoking.

The manuscript is concise enough and appropriately referenced.

The presented results are based on a powerful populations-based cohort in Sweden and Estonia. The results and conclusions in this manuscript are internally consistent.

However, there are three major points that prevent me from drawing final conclusions on the presented results (see below). The first one relates two the study design and the second to the assumed genetic model.

Major Revisions

1. Unfortunately, the description of the study population is too vague.
   The number of n=767 subjects from Sweden and Estonia (the study population) cannot be reproduced from reference 9 given for ECRHS II study design. Furthermore, it cannot be assessed if this population is a fair representation of the randomly drawn cohort.

2. The ECRHS has a complex survey design:
   a) The samples for ECRHS I was a stratified random sample of the participating communities. Stratification was performed for sex. This stratification would best be accounted for by appropriate survey statistics, but it may be accounted for by adjusting or stratifying for sex. The authors have not done this for all presented analyses and they have to at least give reasoning for that.

   b) Two samples were invited for ECRHS II: a random sample of ECRHS I participants and all "symptomatic" ECRHS I participants not included in the random sample. The authors seem to work on the random sample plus the "symptomatic" sample which leads to artificially high prevalence of symptoms. This should be described in detail and the authors should appropriately report or correct for the enrichment of "symptomatics". Estimators such as disease prevalence from the combined (random plus "symptomatic") cohort are not
particularly meaningful and may mislead the reader.

3. There is no mention of the underlying genetic model that has been assumed. The authors seem to have conducted a 2DF Chi-Square test in most instances assuming a "co-dominant" model. It would be good to have indication in the text which models were assumed and tested in this publication and to get reasoning for that either based on the presented data or already published data.

The authors have provided stratified and pooled analyses in particular for study centre, study design (random vs. random + "symptomatic cohort"), and sex. However, this has not been done consistently throughout the manuscript. For some of the pooled estimators it is questionable whether they are meaningful and should be presented in this way.

My suggestion would be to stick to the random cohort. The tables should concentrate on the results from the random cohort rather than the combined cohort if the authors decide to stick to reporting effects for both cohorts.

Analyses should always be stratified for sex or adjusted for sex when the stratification indicates that both sex groups can be pooled. The same is true for study centre.

**Level of interest:** An article of importance in its field

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests.