Reviewer's report

Title: The role of IREB2 and transforming growth factor beta-1 genetic variants in COPD: a replication case-control study

Version: 1 Date: 6 November 2010

Reviewer: Benjamin Raby

Reviewer's report:

This is a well-powered case-control study of 1002 COPD cases and 900 controls, attempting to replicate genetic association of single nucleotide polymorphisms in the IREB (7 SNP) and TGFB1 (4 SNP) with COPD. DNA samples were collected from white Caucasian subjects at six European centers, including Barcelona, Bristol, Dublin, Edinburgh, Leiden and Pisa. COPD cases and control definitions are reasonable.

SNPs previously associated with COPD were selected for typing. The genotyping was carried out commercially by K-Bioscience, though the specific application/assay is not specified. QC included duplicate genotyping of 5% of samples, and replicate genotyping with a second set of platforms, including enzyme digestion, allele-specific PCR and sequencing. The results of these analyses suggest excellent quality of the genotypes. There is no mention of testing for deviations from Hardy Weinberg Equilibrium in the controls, which would be useful to screen for cryptic population substructure. Statistical analysis was performed in SAS, including calculation of LD, and genetic association by logistic regression with adjustment for age, sex, smoking, center, and with interactions for age, sex, and smoking with center. 2 and 3 SNP haplotypes were also assessed using FAMHAP18. Severity testing was performed using GLM and Chi-Square statistics.

Though 11 SNPs were typed, based on high LD between sets of SNPs, data is presented for only 6 of the variants (3 in each gene). Despite very good power, no association was observed for the 3 TGFB SNPs. In contrast, all IREB2 SNP showed significant association with COPD. Haplotype analysis was similar to the single SNP analysis. There was no association with severity.

Overall, this is a well conducted replication study. I suspect that the main results are correct, as the IREB locus has been consistently replicated in association with COPD in several cohorts. There are several issues, however, that the authors should address:

1) Population stratification: The authors need to provide considerably more data to convince the reviewer that the associations are being accurately estimated, without confounding by population admixture. Suggested additional analyses include: (i) Hardy Weinberg Equilibrium estimates in all controls, and then by study center; (ii) reports of the allele frequencies in each center, with tests for differences in allele frequency between centers; (iii) sensitivity analysis of the
case-control study by repeating the analysis in each center separately. For this analysis, it is the odds ratios that are important, not the specific p-values. (iv) Ideal, a test for spurious admixture using a set of random variants (either AIMS markers or random sets) would be very helpful for both this and future studies to exclude admixture.

2) IREB2 vs. genic region: The authors make claims that the associations observe support IREB2 as the COPD susceptibility gene, rather than LOC123688 or CHRNA3. There is insufficient data to support this claim. First, the authors did not test SNP in these other genes. More importantly, prior studies have demonstrated strong associations across the larger region, and it remains unclear which gene is the culprit. Finally, it is possible that several genes here contribute to COPD pathogenesis. The authors should be more conservative and simply conclude that their data supports this chromosomal region as a COPD susceptibility locus, rather than making definitive statements regarding one gene over another.

3) Corrections for multiple comparisons

4) The discussions regarding TGFB1 and MMP12 are speculative. No data is provided regarding MMP12 genotypes, so it is not clear why that gene is discussed as a possible confounder here. There are many reasons for failure of replication. The authors should discuss the most plausible.

5) Limitations: The authors do not review the limitations of their work, many of which are alluded to in this review. First, the issue of stratification. Second, the limited testing of only a handful of SNP, rather than a comprehensive survey of the region. Three, there are likely many unobserved variants that may be contributing to the association, which would only be identified via comprehensive resequencing. Four, the difficulties in making conclusive statements regarding specific genes in the context of genetic association studies where LD is extensive.

**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