



Published in final edited form as:

Thorax. 2011 December ; 66(12): 1085–1090. doi:10.1136/thoraxjnl-2011-200017.

Identification of FGF7 as a novel susceptibility locus for chronic obstructive pulmonary disease

John M. Brehm^{1,2,3,5,7}, Koichi Hagiwara⁹, Y Tesfaigzi¹⁵, S Bruse¹⁵, Thomas J. Mariani¹⁰, Soumyaroop Bhattacharya¹⁰, Nadia Boutaoui¹, John P. Ziniti², Manuel E. Soto-Quiros⁸, Lydiana Avila⁸, Michael H. Cho^{2,3,5}, Blanca Himes², Augusto A. Litonjua^{2,3,5,7}, Francine Jacobson⁶, Per Bakke¹², Amund Gulsvik¹², Wayne H Anderson¹³, David A. Lomas¹⁴, Erick Forno¹¹, Soma Datta², Edwin K. Silverman^{2,3,4,5}, and Juan C. Celedón^{1,2,3,4,5,7}

¹Division of Pediatric Pulmonary Medicine, Allergy, and Immunology, Children's Hospital of Pittsburgh of UPMC, Pittsburgh, PA

²Channing Laboratory, Brigham and Women's Hospital, Boston, Massachusetts

³Division of Pulmonary/Critical Care Medicine, Brigham and Women's Hospital, Boston, Massachusetts

⁴Center for Genomic Medicine, Brigham and Women's Hospital, Boston, Massachusetts

⁵Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts

⁶Department of Radiology, Brigham and Women's Hospital, Boston, Massachusetts

⁷Department of Medicine, Harvard Medical School, Boston, Massachusetts

⁸Division of Pediatric Pulmonology, Hospital Nacional de Niños, San José, Costa Rica

⁹Saitama Medical University Hospital and Institute, Saitama, Japan

¹⁰University of Rochester Medical Center, Rochester, NY

¹¹Division of Pediatric Pulmonology, Dept. of Pediatrics, University of Miami, Miami, FL

¹²Haukeland University Hospital and Institute of Medicine, University of Bergen, Bergen, Norway

¹³GlaxoSmithKline Research and Development, Research Triangle Park, North Carolina, USA

¹⁴Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK

¹⁵Lovelace Respiratory Research Institute, Albuquerque, NM

Abstract

Rationale—Traditional genome-wide association studies (GWAS) of large cohort of subjects with chronic obstructive pulmonary disease (COPD) have successfully identified novel candidate genes, but several other plausible loci do not meet strict criteria for genome-wide significance after correction for multiple testing.

Objectives—We hypothesize that by applying unbiased weights derived from unique populations we can identify additional COPD susceptibility loci.

Methods—We performed a homozygosity haplotype analysis on a group of subjects with and without COPD to identify regions of conserved homozygosity (RCHH). Weights were constructed

based on the frequency of these RCHH in case vs. controls, and used to adjust the P values from a large collaborative GWAS of COPD.

Results—We identified 2,318 regions of conserved homozygosity, of which 576 were significantly ($P < .05$) overrepresented in cases. After applying the weights constructed from these regions to a collaborative GWAS of COPD, we identified two single nucleotide polymorphisms in a novel gene (*FGF7*) that gained genome-wide significance by the false discovery rate method. In a follow-up analysis, both SNPs (rs12591300 and rs4480740) were significantly associated with COPD in an independent population (combined P values of 7.9E-07 and 2.8E-06 respectively). In another independent population, increased lung tissue *FGF7* expression was associated with worse measures of lung function.

Conclusion—Weights constructed from a homozygosity haplotype analysis of an isolated population successfully identify novel genetic associations from a GWAS on a separate population. This method can be used to identify promising candidate genes that fail to meet strict correction for multiple testing.

INTRODUCTION

Traditional genome-wide association studies (GWAS) have identified novel susceptibility loci for complex diseases such as chronic obstructive pulmonary disease (COPD) ^{1–3}. Because the effect size of most common disease-susceptibility variants is modest, GWAS of complex diseases require large sample sizes to achieve statistically significant results after correction for multiple testing. Weighting the results of GWAS according to prior information (e.g., from linkage studies) may significantly improve power to detect associations that do not meet genome-wide significance ⁴.

Homozygosity mapping is a promising technique to identifying regions of the genome that are more likely to contain disease susceptibility loci. Although initially developed to identify rare susceptibility mutations for monogenic traits in families ⁵, homozygosity mapping has recently been successfully applied to the study of complex diseases ^{6,7}. While techniques vary, the concept underlying all homozygosity haplotype methods is that regions of homozygosity are more likely to contain disease-susceptibility loci in affected subjects than in unaffected individuals⁸.

Using high-density SNP arrays, Miyazawa et. al. developed a novel variation on homozygosity mapping that tests whether multiple subjects share the same genotype among homozygous SNPs, and then constructs a region of conserved homozygosity haplotype that reflects the transmission of the haplotype from a founder population. In theoretical simulations, this method was shown to be a viable method to detect disease-susceptibility loci in recently admixed populations ⁹. We hypothesized that application of this method to a genetic isolate in Costa Rica would result in detection of an over-representation of regions of conserved homozygosity in subjects affected with COPD compared with unaffected subjects. In this report, we first identify regions of conserved homozygosity in Costa Ricans and then show that weights derived from these regions can be applied to GWAS in non-isolated populations to identify novel disease-susceptibility loci for COPD. Using this approach we identify a novel COPD candidate gene (*FGF7*).

MATERIALS and METHODS

Study Population

The primary study population consisted of 58 subjects with (cases) and 57 subjects without (controls) COPD in the Genetic Epidemiology of COPD in Costa Rica (GECCOR) study. Cases were recruited from patients attending four adult hospitals in San Jos  (Costa Rica)

and their affiliated clinics, and through newspaper ads. Control subjects were recruited from individuals attending a smoking-cessation clinic at the Institute for Pharmaco-dependency in San Jos , and through newspaper ads. To ensure their descent from the founder population of the Central Valley of Costa Rica (which is predominantly of Spanish and Native American ancestry), all participants were required to have at least six great-grandparents born in the Central Valley. Additional inclusion criteria for cases were ages 21 to 71 years, physician-diagnosed COPD, 10 pack-years of cigarette smoking, and an FEV1 65% predicted and an FEV1/FVC ratio of 70% after bronchodilator administration (180 mcg of albuterol by metered dose inhaler). Controls were recruited on the basis of the same criteria for age and smoking history, but they had to have no physician-diagnosed COPD and normal spirometry. Exclusion criteria for cases and controls included history of chronic pre-existing chronic lung disease (e.g., bronchiectasis) and severe alpha-1-antitrypsin deficiency (for cases), based on molecular phenotyping. The baseline characteristics of this cohort are in supplementary table 2.

Written consent was obtained from participating subjects. The study was approved by the Institutional Review Boards of the Hospital Nacional de Ni os (San Jos , Costa Rica), Partners Healthcare System (Boston, MA), and participating NETT, ECLIPSE and Norway centers.

Genotyping of Costa Rican cohort

High-density SNP genotyping was performed using the Illumina Quad 610 platform at the Channing Laboratory, Boston, MA. Cases and controls were randomly distributed among batches, and each batch contained a replicate sample. All subjects had a SNP call rate > 95%. After quality control measures (see supplementary table 1 in the Online Supplement), a total of 558,929 SNPs were acceptable for analysis.

Collaborative COPD Cohorts for Primary GWAS

Three populations with a total of 2,940 cases and 1,380 controls were used for the primary GWAS: (i) subjects in a case-control study of COPD in Norway (838 cases and 791 controls)³, (ii) subjects in the National Emphysema Treatment Trial (NETT, 366 cases) and the Normative Aging Study (NAS, 414 controls)¹⁰¹¹ and (iii) 1,736 cases and 175 controls from the multicenter Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE) Study¹². All controls were current or former smokers with normal spirometry, and all cases with COPD had moderate to very severe disease according to the Global Initiative for Chronic Obstructive Lung Disease (GOLD) classification¹³.

Lovelace Smokers Cohort

The top SNPs in novel genes were replicated in a cohort of 1,845 smoking adults in New Mexico, 424 (23%) of whom were classified with COPD based on FEV1/FVC ratio below the 5th percentile of the predicted value, also referred to as the lower limit of normal (LLN)¹⁴. Of the 1,845 participants, 1,411 (77%) were white and 313 (17%) were Hispanic. The protocols for subject recruitment and data collection for the Lovelace Smokers Cohort has been previously described in detail.¹⁵ The two SNPs (rs12591300 and rs4480740) were genotyped by allelic discrimination using Taqman assay (Applied Biosystems, Foster City, CA). The case-control association analysis was first performed in all subjects, and then separately in whites and Hispanics. All analyses were adjusted for age, gender, and pack-years of cigarette smoking; the analysis of all subjects was additionally adjusted for self-declared ethnicity.

Gene Expression Analysis

For the top novel candidate genes, we examined the correlation of gene expression in lung tissue with COPD intermediate phenotypes (FEV1 and FEV1/FVC ratio) in a previously published COPD biomarker discovery study¹⁶. This cohort consists of 56 subjects with varying degrees of obstruction who underwent lung resection for a solitary pulmonary nodule. RNA expression profiling was completed using the Affymetrix U133 Plus 2.0 array, as previously described¹⁶. Expression correlation with quantitative phenotypes was conducted as previously described¹⁶.

Statistical Analysis

Construction of RCHH—Regions of conserved homozygosity haplotype (RCHHs) were identified using the method described by Miyazawa et. al.⁹. In brief, for any given individual all heterozygous SNPs were ignored and the SNP location was scored with the value of the allele for that subject. Subjects are compared only across SNPs that are scored. RCHH are defined by runs of SNPs that share the same allele at the homozygous locations across multiple subjects, ignoring heterozygous SNPs. The size of the shared segments between any two individuals was set at 3.0 cM (roughly and approximately 3 million base pairs), which in theoretical work conducted by Miyazawa et. al.⁹ reduced the false positive and false negative rates of discovery. A theoretical ancestral segment is then constructed from the largest subgroup of subjects sharing a particular RCHH (see Supplementary Figure 1). While any two subjects must have at least 3.0 cM of sharing, the size may be much smaller when comparing across multiple subjects (Supplementary figure 2). If more than one ancestral region is identified at a particular chromosomal location, the region shared by the most number of subjects is used (Supplementary figure 3). The total number of cases and controls sharing this ancestral allele is used to calculate a P value based on a standard normal distribution.

For the primary analysis of the Collaborative COPD cohort, logistic regression analysis was performed under an additive genetic model for each SNP, adjusting for age, pack-years of smoking, and the first 16 principal components (to adjust for population stratification). The P values from all RCHH identified in Costa Rica were then used to construct a cumulative weight for each SNP from the recent GWAS of COPD in the combined cohort of Norway, ECLIPSE, and NETT-NAS using the method developed by Roeder et. al.⁴ Briefly, the weighting method utilizes prior information (in this case the P value representing the degree of overrepresentation of a region of the genome in cases vs. controls) to upweight or downweight P values from an association study (in this case the GWAS of COPD in the collaborative cohort). In order to maintain an overall alpha level of 0.05, the assigned weights across the genome average to 1. For this study, SNPs that did not fall inside of an RCHH (and therefore did not have a P value) were assigned a P value of 1 (and therefore a weight approaching zero). This is a more conservative approach than excluding these SNPs from consideration. The method then calculates a false-discovery rate (FDR) using the method described by Benjamini and Hochberg¹⁷ to correct for multiple testing.

The RCHH were created and compared with HAnalysis (available at <http://www.hhanalysis.com>). Association analysis was performed using PLINK v.1.07 (<http://pngu.mgh.harvard.edu/purcell/plink>). The weighting procedure was performed using software developed by Roeder et al.⁴ (<http://wpicr.wpic.pitt.edu/wpiccompngen/>). All other statistical analysis was performed using R version 2.9.0 (<http://www.R-project.org>).

RESULTS

Identification of RCHH in Costa Rica and Construction of Weights

In total, 2,318 regions of conserved homozygosity haplotype (RCHH) were identified in the Costa Rica cohort. Of these 2,318 regions, 576 were significantly ($P < .05$) overrepresented in cases compared to controls; none of the regions were significantly more frequent in controls than cases. The median size of the significant regions was 105 kb, and the largest was 7.2 Mb. Supplementary table 3 in the Online Supplement shows the top 20 P values representing 100 RCHH in Costa Rica.

Each SNP in the combined collaborative COPD cohort was then mapped to a region of conserved homozygosity haplotype and assigned the P value of the whole region. SNPs that did not map to an RCHH were assigned a P value of 1. The mapped P values across all genotyped SNPs were then used to create weights using a cumulative distribution function. The algorithm is constructed so that the mean weight across all SNPs is 1: some SNPs are upweighted and a much larger fraction are downweighted. The nominal P value is divided by the weight to obtain the weighted P value.

Application of Weights to COPD GWAS

We applied the weights derived from the homozygosity haplotype analysis above to reanalyze GW genotypic data in a cohort of subjects of European descent that was previously employed for a traditional GWAS of COPD. After weighting, 14 SNPs were significant at an FDR-corrected alpha of .05. The top 5 SNPs from the unweighted GWAS retained their original ranks, but several SNPs that did not achieve genome-wide significance in the traditional GWA analysis became more statistically significant and moved higher in the list (Table 1). Of these SNPs, those in the gene for *FAM13A* were identified in the original analysis of the GWAS¹, and SNPs in *IREB2*¹⁸, and *CHRNA3*³ have been implicated in COPD affection status in prior candidate-gene and GW-association studies. Two of the other SNPs lie in two novel candidate genes for COPD: fibroblast growth factor-7 (*FGF7*) and proteasome subunit, alpha-type, 4 (*PSMA4*) (Figure 1). The RCHH in Costa Rica that contains *FGF7* was present in 7 cases and no controls, and the RCHH containing *PSMA4* was present in 5 cases and no controls.

The regions containing the genes *CHRNA3* and *IREB2* were also over-represented in cases compared to controls ($P < .05$), and after weighting they were genome-wide significant by FDR. While there was an RCHH containing *FAM13A* identified in the Costa Rica cohort, it was only seen in one case and no controls.

Replication in Lovelace Smokers Cohort

The top two SNPs in or near *FGF7* were genotyped in the 1,845 smoking adults in the Lovelace Smokers Cohort. The minor alleles of both SNPs conferred increased odds for COPD in the whole population in the same direction as the original collaborative COPD cohort (Table 2). Among the Hispanic subgroup, the effect size was larger and in the same direction for both SNPs, but only rs12591300 showed a significant association with COPD affection status.

Gene Expression Analysis

Our previous studies indicate that gene expression patterns associated with quantitative, intermediate COPD phenotypes are most informative for the discovery of disease-associated genes^{16,18,19}. We examined disease-associated expression patterns for our novel candidate genes in a previously published genome-wide expression data set from 56 subjects with varying degrees of airflow obstruction (assessed by spirometric measures of lung function

[FEV1 and FEV1/FVC])¹⁶. Expression of *FGF7* (as defined by multiple and independent probe sets) was significantly negatively correlated with both FEV1 (nominal P value < 0.01) and FEV1/FVC (nominal P value < 0.01), indicating increased expression associated with increased disease severity. Expression in COPD subjects was increased compared to control subjects, but the difference was not statistically significant. *PSMA4* expression was not correlated with lung function, and was not differentially expressed in cases vs. controls.

DISCUSSION

While successful in identifying novel candidate genes, GWAS of complex traits are unlikely to identify all potential common disease-susceptibility variants because of limited power if strict criteria for GW significance are applied. In the absence of a very large sample size, novel methods are needed to identify disease-susceptibility variants not meeting GW significance. We identified RCHH for COPD in a GW case-control study in Costa Rica. After applying a weighting method based on the degree of significance of these regions to a GWAS of COPD cases and controls of European descent, we identified two SNPs in a novel candidate gene for COPD (*FGF7*) and demonstrated that several SNPs in the previously identified candidate genes *IREB2* and *CHRNA3* met GW criteria for statistical significance. A SNP in another novel gene (*PSMA4*) was GW significant after weighting. However, expression of *PSM4* in the lung was not associated with COPD phenotypes, and thus the observed association is likely due to LD with the nearby genes *CHRNA3* and *IREB2*. We then replicated the two *FGF7* SNPs in an independent cohort of smoking adults, and showed that they are both significantly associated in the same direction with COPD. Notably, the effect sizes in Hispanics are larger than in the overall cohort, suggesting that these alleles confer greater risk in this population. This Hispanic population in New Mexico has a similar proportion of European and Native American ancestry as the Costa Rica cohort²⁰²¹, so another likely possibility is that patterns of linkage disequilibrium (LD) may be different between Hispanics and Caucasians in this genomic region, and that these SNPs are tagging a haplotype or functional SNP in the Hispanic subjects. Additionally, there was a trend towards increased lung tissue expression of *FGF7* in an independent cohort of COPD subjects, in whom there was a significant negative correlation between *FGF7* expression and FEV1 and FEV/FVC ratio.

There are several plausible methods for weighting chromosomal regions in GWAS, including upweighting previously identified candidate genes, coding variants, exons, and promoter regions. However, these weighting strategies work counter to one of the strengths of a GWAS: its hypothesis-free nature. Using homozygosity haplotypes as a weighting method avoids the pitfall of these other weighting strategies because they are constructed using a hypothesis-free method, so the weights are unbiased with respect to prior knowledge.

One of the main strengths of our study is that it shows the power of using HH analysis in an isolated population to investigate common diseases. While our sample size was small, Mizayawa et al. have previously shown in simulated data that HH analysis has the ability to identify the region containing a SNP inherited identity-by-descent from a distant common ancestor using only 45 cases and 45 controls. In our own data, we were able to show that the previously identified candidate genes *IREB2* and *CHRNA3* fall within an RCHH that is significantly overrepresented in subjects with COPD. When combined with results from a weighted GWAS in an independent cohort with adequate sample size, we were able to show that variants in these genes are significant after correction for multiple testing.

Two novel genes are contained within significant regions of conserved homozygosity, and after weighting they are significant by FDR correction. The first, *FGF7* or fibroblast growth

factor 7, was identified in cultured human embryonic lung fibroblasts²², and plays a role in promoting wound healing²³ and protecting airway epithelium from oxidant injury in mice²⁴. One of the SNPs identified in this study (rs4480740) is in an intron of *FGF7*, and the other (rs12591300) lies immediately upstream of *FGF7* in an intron of hypothetical protein LOC196951. In a GWAS of FEV1 in the British 1958 Birth Cohort, 5 out of the 9 SNPs genotyped in *FGF7* were significantly ($P < .05$) associated with differences in lung function, although not the two SNPs identified in this study.²⁵ *FGF7* has been shown to protect against oxidative stress response specifically in the lung epithelium²⁴, so increases in expression associated with disease progression may indicate a greater burden of injury. A limitation of our study is lack of experimental evidence for an effect(s) of the SNPs identified in *FGF7* on gene expression. We hypothesize that these SNPs cause decreased expression of *FGF7*, which could affect antioxidant mechanisms protecting against detrimental effects of cigarette smoking on the lung. Alternatively, *FGF7* may play a role in disease susceptibility through its role in epithelial development during embryogenesis by influencing epithelial responses to cigarette smoke. Since it is unclear whether increased *FGF7* expression is a marker of exposure to oxidant injury or a cause of epithelial damage, further work must be done to characterize the role of these SNPs on *FGF7* expression.

The HHAnalysis algorithm works best under certain assumptions, namely that 1) the risk alleles were introduced into the population from a population of common ancestors within the last several hundred years, 2) the target population is genetically isolated, 3) the number of common ancestors introducing the risk allele is small, and 4) that the risk of the disease allele is moderate to high. Violations of these assumptions reduce the theoretical expected size of the RCHH and/or the association of the RCHH with disease, which reduces the power of the algorithm to detect them. Genetic and historical data for the population of the Central Valley of Costa Rica suggests that the first three assumptions are met. As in most association studies of complex disease, the effect size of a risk allele is likely small to moderate at most, and we expect that this has somewhat reduced our power.

Whereas other homozygosity mapping methods are primarily designed to detect recessive alleles, the HHAnalysis method instead uses homozygosity to identify ancestral regions inherited from a common ancestor. These RCAs can harbor risk alleles that operate under recessive, dominant, or additive models. However, the HHAnalysis algorithm would also detect copy number variation that results in the deletion of a single allele. While this may explain a fraction of the regions identified, the top novel SNPs identified in *FGF7* do not fall within known regions of copy number variation according to the Database of Genomic Variants.²⁶

In summary, we have shown that weights obtained from HH analysis in an isolated population can improve the power to detect novel variants in GWAS in non-isolates. In addition to confirming results for previously identified variants in *IREB2* and *CHRNA3*, we have identified variants in a novel candidate gene (*FGF7*) for COPD. The validity of this gene is supported by replication in an independent cohort of smoking adults, and expression data showing consistent and significant patterns associated with COPD intermediate lung function phenotypes. Further analysis of these genes in the Costa Rica cohort and functional studies should yield insights into the causative SNPs or haplotypes that underlie the associations identified in this study.

What is the key question? Can information from isolated populations improve our ability to detect novel genetic variants in genome-wide association studies (GWAS)?

What is the bottom line? We identified statistically significant polymorphisms in a novel COPD gene (*FGF7*), which we replicated in an independent population.

Why read on? We demonstrate the use of a novel method (Homozygosity Haplotype Analysis) for identifying genomic regions that are inherited from a common ancestor, and use that information to weight a GWAS of COPD to identify novel genetic variants that are associated with increased risk of disease.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding/Support: The Genetic Epidemiology of COPD in Costa Rica is supported by grant R01HL073373 from the National Heart, Lung and Blood Institute. The National Emphysema Treatment Trial (NETT) is supported by contracts with the National Heart, Lung, and Blood Institute (N01HR76101, N01HR76102, N01HR76103, N01HR76104, N01HR76105, N01HR76106, N01HR76107, N01HR76108, N01HR76109, N01HR76110, N01HR76111, N01HR76112, N01HR76113, N01HR76114, N01HR76115, N01HR76116, N01HR76118, N01HR76119), the Centers for Medicare and Medicaid Services (CMS); and the Agency for Healthcare Research and Quality (AHRQ). The Norway cohort and the ECLIPSE study (Clinicaltrials.gov identifier NCT00292552; GSK Code SCO104960) are funded by GlaxoSmithKline. The Lovelace Smokers Cohort is supported by funding from the State of New Mexico (appropriation from the Tobacco Settlement Fund) and by grant RO1 ES015482 from the National Institute of Environmental Health Sciences. We acknowledge use of genotype data from the British 1958 Birth Cohort DNA collection, funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02.

References

1. Cho MH, Boutaoui N, Klanderman BJ, Sylvia JS, Ziniti JP, Hersh CP, et al. Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat Genet.* 2010; 42(3):200–202. [PubMed: 20173748]
2. Hirschhorn JN. Genomewide association studies—illuminating biologic pathways. *N Engl J Med.* 2009; 360(17):1699–1701. [PubMed: 19369661]
3. Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, et al. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet.* 2009; 5(3) e1000421.
4. Roeder K, Bacanu SA, Wasserman L, Devlin B. Using linkage genome scans to improve power of association in genome scans. *Am J Hum Genet.* 2006; 78(2):243–252. [PubMed: 16400608]
5. Lander ES, Botstein D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science.* 1987; 236(4808):1567–1570. [PubMed: 2884728]
6. Lencz T, Lambert C, DeRosse P, Burdick KE, Morgan TV, Kane JM, et al. Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proceedings of the National Academy of Sciences.* 2007; 104(50):19942.
7. Morrow EM, Yoo SY, Flavell SW, Kim TK, Lin Y, Hill RS, et al. Identifying Autism Loci and Genes by Tracing Recent Shared Ancestry. *Science.* 2008; 321(5886):218. [PubMed: 18621663]
8. Reich DE, Schaffner SF, Daly MJ, McVean G, Mullikin JC, Higgins JM, et al. Human genome sequence variation and the influence of gene history, mutation and recombination. *Nat Genet.* 2002; 32(1):135–142. [PubMed: 12161752]
9. Miyazawa H, Kato M, Awata T, Kohda M, Iwasa H, Koyama N, et al. Homozygosity Haplotype Allows a Genomewide Search for the Autosomal Segments Shared among Patients. *The American Journal of Human Genetics.* 2007; 80(6):1090–1102.
10. Bell B, Rose CL, Damon A. The Normative Aging Study: an interdisciplinary and longitudinal study of health and aging. *Aging Hum Dev.* 1972; 3:5–17.
11. Fishman A, Martinez F, Naunheim K, Piantadosi S, Wise R, Ries A, et al. A randomized trial comparing lung-volume-reduction surgery with medical therapy for severe emphysema. *N Engl J Med.* 2003; 348(21):2059–2073. [PubMed: 12759479]

12. Vestbo J, Anderson W, Coxson HO, Crim C, Dawber F, Edwards L, et al. Evaluation of COPD Longitudinally to Identify Predictive Surrogate End-points (ECLIPSE). *Eur Respir J*. 2008; 31(4): 869–873. [PubMed: 18216052]
13. Pauwels RA, Buist AS, Calverley PM, Jenkins CR, Hurd SS. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease. NHLBI/WHO Global Initiative for Chronic Obstructive Lung Disease (GOLD) Workshop summary. *Am J Respir Crit Care Med*. 2001; 163(5):1256–1276. [PubMed: 11316667]
14. Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R, et al. Interpretative strategies for lung function tests. *Eur Respir J*. 2005; 26(5):948–968. [PubMed: 16264058]
15. Sood A, Stidley CA, Picchi MA, Celedon JC, Gilliland F, Crowell RE, et al. Difference in airflow obstruction between Hispanic and non-Hispanic White female smokers. *COPD*. 2008; 5(5):274–281. [PubMed: 18972275]
16. Bhattacharya S, Srisuma S, Demeo DL, Shapiro SD, Bueno R, Silverman EK, et al. Molecular biomarkers for quantitative and discrete COPD phenotypes. *Am J Respir Cell Mol Biol*. 2009; 40(3):359–367. [PubMed: 18849563]
17. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1995; 57(1):289–300.
18. DeMeo DL, Mariani T, Bhattacharya S, Srisuma S, Lange C, Litonjua A, et al. Integration of genomic and genetic approaches implicates IREB2 as a COPD susceptibility gene. *Am J Hum Genet*. 2009; 85(4):493–502. [PubMed: 19800047]
19. Demeo DL, Mariani TJ, Lange C, Srisuma S, Litonjua AA, Celedon JC, et al. The SERPINE2 gene is associated with chronic obstructive pulmonary disease. *Am J Hum Genet*. 2006; 78(2):253–264. [PubMed: 16358219]
20. Klimentidis YC, Miller GF, Shriver MD. Genetic admixture, self-reported ethnicity, self-estimated admixture, and skin pigmentation among Hispanics and Native Americans. *Am J Phys Anthropol*. 2009; 138(4):375–383. [PubMed: 18951390]
21. Morera B, Barrantes R, Marin-Rojas R. Gene admixture in the Costa Rican population. *Ann Hum Genet*. 2003; 67(Pt 1):71–80. [PubMed: 12556237]
22. Rubin JS, Osada H, Finch PW, Taylor WG, Rudikoff S, Aaronson SA. Purification and characterization of a newly identified growth factor specific for epithelial cells. *Proc Natl Acad Sci U S A*. 1989; 86(3):802–806. [PubMed: 2915979]
23. Werner S, Smola H, Liao X, Longaker MT, Krieg T, Hofschneider PH, et al. The function of KGF in morphogenesis of epithelium and reepithelialization of wounds. *Science*. 1994; 266(5186):819–822. [PubMed: 7973639]
24. Ray P, Devaux Y, Stolz DB, Yarlagadda M, Watkins SC, Lu Y, et al. Inducible expression of keratinocyte growth factor (KGF) in mice inhibits lung epithelial cell death induced by hyperoxia. *Proc Natl Acad Sci U S A*. 2003; 100(10):6098–6103. [PubMed: 12732722]
25. [accessed 8/1/2010] <http://www.b58cgene.sgul.ac.uk/>
26. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, et al. Detection of large-scale variation in the human genome. *Nat Genet*. 2004; 36(9):949–951. [PubMed: 15286789]

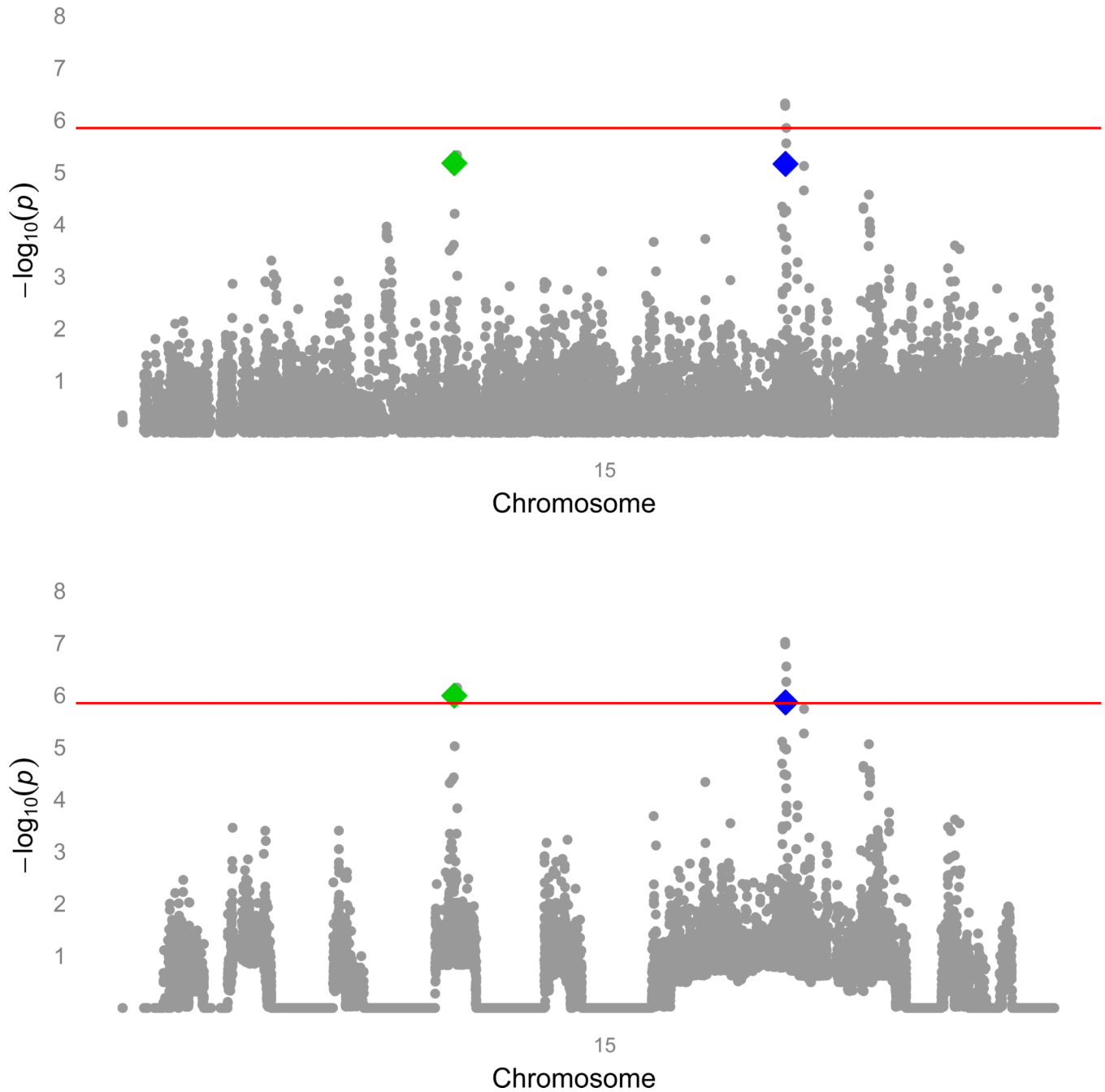


Figure 1. Manhattan plot of chromosome 15, before (top) and after weighting. rs4480740 (green) is in the gene *FGF7*, and rs2036534 (blue) is in the promoter of *PSMA4*. The red line indicates the FDR corrected alpha level for genome-wide significance.

TABLE 1

FDR significantⁱ results from weighted GWAS

SNP	Location (Chr: BP in hg18)	A1	OR	Original P value ⁱⁱ	Original rank	Weighted P value ⁱⁱⁱ	Gene (distance from gene)	Norway P value ^{iv}	NETT-NAS P value	Eclipse P value	Costa Rica RCHH		
											Cases (%)	Controls (%)	RCHH P value
rs1903003	4:9010320	C	0.75	7.18E-08	1	6.87E-08	FAM13A0	4.3E-04	1.4E-03	9.1E-03	1 (2%)	0 (0%)	0.2126
rs7671167	4:90103002	C	0.76	8.59E-08	2	8.22E-08	FAM13A(0)	7.9E-04	2.7E-04	7.8E-03	1 (2%)	0 (0%)	0.2126
rs1062980	15:76579582	C	0.76	4.81E-07	3	9.53E-08	IREB2(0)	9.9E-03	1.0E-02	3.6E-02	5 (9%)	0 (0%)	0.0164
rs13180	15:76576543	C	0.76	5.01E-07	4	9.93E-08	IREB2(0)	7.9E-03	1.6E-02	4.3E-02	5 (9%)	0 (0%)	0.0164
rs8034191	15:76593078	C	1.32	5.37E-07	5	1.06E-07	IREB2(+12.22kb)	1.5E-04	8.7E-03	8.2E-01	5 (9%)	0 (0%)	0.0164
rs12914385	15:76685778	T	1.29	1.42E-06	9	2.81E-07	CHRNA3(0)	1.4E-03	9.8E-03	9.5E-01	5 (9%)	0 (0%)	0.0164
rs1051730	15:76681394	A	1.29	2.80E-06	14	5.54E-07	CHRNA3(0)	4.3E-04	2.1E-02	8.4E-01	5 (9%)	0 (0%)	0.0164
rs17404727	15:47791375	C	1.28	4.71E-06	15	7.17E-07	.	1.9E-02	2.2E-02	3.4E-02	7 (12%)	0 (0%)	0.0049
rs996414	9:26570067	G	0.76	1.80E-06	11	8.76E-07	.	6.2E-04	8.8E-01	2.6E-01	2 (3%)	0 (0%)	0.1063
rs4480740	15:47543134	A	1.27	6.75E-06	17	1.03E-06	FGF7(0)	4.0E-02	2.5E-02	2.1E-02	7 (12%)	0 (0%)	0.0049
rs12591300	15:47492033	A	1.27	8.78E-06	21	1.34E-06	FGF7(-10.72kb)	3.9E-02	8.3E-02	2.5E-02	7 (12%)	0 (0%)	0.0049
rs2656069	15:76532762	C	0.75	6.82E-06	18	1.35E-06	IREB2(0)	1.6E-01	2.6E-03	1.0E-02	5 (9%)	0 (0%)	0.0164
rs2036534	15:76614003	C	0.75	6.98E-06	19	1.38E-06	PSMA4(-5.798kb)	5.8E-02	7.4E-03	9.4E-02	5 (9%)	0 (0%)	0.0164
rs2869967	4:90088355	C	1.29	1.48E-06	10	1.41E-06	FAM13A1(0)	4.7E-04	7.6E-03	4.4E-03	1 (2%)	0 (0%)	0.2126

ⁱAn FDR corrected P value of 1.43×10^{-6} was used as the cutoff for genome-wide significance.

ⁱⁱResults previously published by Cho, et. al. ¹

ⁱⁱⁱThe weighted P value is the original P value divided by the weight constructed from the RCHH (not shown).

^{iv}P values for individual cohorts are the original, unweighted P values.

TABLE 2

Combined P values for replication of FGF7 SNPs

SNP	COPD consortium OR (2-sided P value)	Lovelace (All subjects) OR (1-sided P value)	Lovelace (Hispanics) OR (1-sided P value)	Combined 1-sided P value (All subjects)	Combined 1-sided P value (Hispanics only) ^V
rs12591300	1.27 (8.78E-06)	1.21 (.01)	1.5 (.06)	7.9E-07	3.9E-06
rs4480740	1.27 (6.75E-06)	1.15 (0.05)	1.64 (.025)	2.8E-06	1.5E-06

^V Fisher's combined P value using original 2-sided p values.