Effect of Five Genetic Variants Associated with Lung Function on the Risk of Chronic Obstructive Lung Disease, and Their Joint Effects on Lung Function

María Soler Artigas[1], Louise V. Wain[1], Emmanouela Repapi[1,2], Ma'en Obeidat[3], Ian Sayers[3], Paul R. Burton[1], Toby Johnson[4], Jing Hua Zhao[5], Eva Albrecht[6], Anna F. Dominiczak[7], Shona M. Kerr[8], Blair H. Smith[9], Gemma Cadby[10,11], Jennie Hui[12–14], Lyle J. Palmer[10,11], Aroon D. Hingorani[15], S. Goya Wannamethee[16], Peter H. Whincup[17], Shah Ebrahim[18], George Davey Smith[19], Inês Barroso[20,21], Ruth J. F. Loos[5], Nicholas J. Wareham[5], Cyrus Cooper[22], Elaine Dennison[22], Seif O. Shaheen[23], Jason Z. Liu[24], Jonathan Marchini[24], Medical Research Council National Survey of Health and Development (NSHD) Respiratory Study Team[25,26*], Santosh Dahgam[27], Åsa Torinsson Naluai[28], Anna-Carin Olin[27], Stefan Karrasch[29], Joachim Heinrich[30], Holger Schulz[30], Tricia M. McKeever[31,32], Ian D. Pavord[33], Markku Heliövaara[34], Samuli Ripatti[34,35], Ida Surakka[34,35], John D. Blakey[3], Mika Kähönen[36], John R. Britton[31,32], Fredrik Nyberg[27,37], John W. Holloway[38,39], Debbie A. Lawlor[19], Richard W. Morris[16], Alan L. James[13,40], Cathy M. Jackson[41], Ian P. Hall[3], Martin D. Tobin[1], and the SpiroMeta Consortium[†]

[1]Departments of Health Sciences and Genetics, University of Leicester, Leicester; [2]Ludwig Institute for Cancer Research, University of Oxford, Oxford, United Kingdom; [3]Division of Therapeutics and Molecular Medicine, Nottingham Respiratory Biomedical Research Unit, University Hospital of Nottingham, Nottingham, United Kingdom; [4]Clinical Pharmacology, William Harvey Research Institute, Barts and London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom; [5]Medical Research Council (MRC) Epidemiology Unit, Institute of Metabolic Science, Cambridge, United Kingdom; [6]Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; [7]College of Medical, Veterinary, and Life Sciences, University of Glasgow, Glasgow, Scotland; [8]Medical Genetics, University of Edinburgh, Molecular Medicine Centre, Edinburgh, Scotland; [9]Centre of Academic Primary Care, University of Aberdeen, Scotland, United Kingdom; [10]Ontario Institute for Cancer Research and [11]Samuel Lunenfeld Research Institute, Toronto, Ontario, Canada; [12]Molecular Genetics, PathWest Laboratory Medicine, Nedlands, Western Australia; [13]Busselton Population Medical Research Foundation, Sir Charles Gairdner Hospital, Nedlands, Western Australia; [14]Schools of Population Health and Pathology and Laboratory Medicine, University of Western Australia, Perth, Western Australia, Australia; [15]Department of Epidemiology and Public Health, University College London, London, United Kingdom; [16]Department of Primary Care and Population Health, University College London, London, United Kingdom; [17]Division of Community Health Sciences, St. George's University of London, London, United Kingdom; [18]Non-communicable Diseases Epidemiology Unit, Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom; [19]MRC Centre for Causal Analyses in Translational Epidemiology, School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom; [20]Wellcome Trust Sanger Institute, Cambridge, United Kingdom; [21]University of Cambridge Metabolic Research Laboratories, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, United Kingdom; [22]MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, United Kingdom; [23]Centre for Health Sciences, Barts and London School of Medicine, London, United Kingdom; [24]Department of Statistics, University of Oxford, Oxford, United Kingdom; [25]MRC National Survey of Health and Development, MRC Unit for Lifelong Health and Ageing, London, United Kingdom; [26]MRC-HPA (Health Protection Agency) Centre for Environment and Health, Imperial College London, St. Mary's Campus, London, United Kingdom; [27]Occupational and Environmental Medicine and [28]Department of Microbiology and Immunology, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; [29]Institute and Outpatient Clinic for Occupational, Social, and Environmental Medicine, Ludwig Maximilian University, Munich, Germany; [30]Institute of Epidemiology I, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; [31]Division of Epidemiology and Public Health, School of Community Health Sciences, University of Nottingham, City Hospital and [32]Nottingham Respiratory Biomedical Research Unit, University of Nottingham, Nottingham, United Kingdom; [33]Institute for Lung Health, Glenfield Hospital, University Hospitals of Leicester National Health Service Trust, Leicester, United Kingdom; [34]National Institute for Health and Welfare, Helsinki, Finland; [35]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland; and; [36]Department of Clinical Physiology, University of Tampere and Tampere University Hospital, Tampere, Finland; [37]AstraZeneca Research and Development, Mölndal, Sweden; [38]Human Genetics Division and [39]Infection, Inflammation, and Immunity Division, School of Medicine, University of Southampton, Southampton General Hospital, Southampton, United Kingdom; [40]Department of Pulmonary Physiology/West Australian Sleep Disorders Research Institute, Sir Charles Gairdner Hospital, Nedlands, Western Australia, Australia; and [41]University of St. Andrews, St. Andrews, Scotland, United Kingdom

(*Received in original form February 1, 2011; accepted in final form May 26, 2011*)

Supported/partially supported through the European Community's Seventh Framework Programme (FP7/2007-2013), ENGAGE project, grant agreement HEALTH-F4-2007-201413. Complete information concerning funding may be found before the beginning of the REFERENCES.

* See the online supplement for MRC National Survey of Health and Development (NSHD) Respiratory Study Team membership list.

† A list of the SpiroMeta Consortium membership is available in the online supplement.

Am J Respir Crit Care Med Vol 184. pp 786–795, 2011
Originally Published in Press as DOI: 10.1164/rccm.201102-0192OC on June 16, 2011
Internet address: www.atsjournals.org

Complete information concerning author contributions may be found before the beginning of the REFERENCES.

Correspondence and requests for reprints should be addressed to Martin D. Tobin, M.B.Ch.B., MRC Senior Clinical Fellow, Genetic Epidemiology Group, Departments of Health Sciences and Genetics, 2nd Floor, Adrian Building, University of Leicester, University Road, Leicester LE1 7RH, UK. E-mail: mt47@le.ac.uk

This article has an online supplement, which is available from this issue's table of contents at www.atsjournals.org

***Rationale*: Genomic loci are associated with FEV_1 or the ratio of FEV_1 to FVC in population samples, but their association with chronic obstructive pulmonary disease (COPD) has not yet been proven, nor have their combined effects on lung function and COPD been studied.**

***Objectives*: To test association with COPD of variants at five loci (*TNS1*, *GSTCD*, *HTR4*, *AGER*, and *THSD4*) and to evaluate joint effects on lung function and COPD of these single-nucleotide polymorphisms (SNPs), and variants at the previously reported locus near *HHIP*.**

***Methods*: By sampling from 12 population-based studies (n = 31,422), we obtained genotype data on 3,284 COPD case subjects and 17,538 control subjects for sentinel SNPs in *TNS1*, *GSTCD*, *HTR4*, *AGER*, and *THSD4*. In 24,648 individuals (including 2,890 COPD case subjects and 13,862 control subjects), we additionally obtained genotypes for rs12504628 near *HHIP*. Each allele associated with lung function decline at these six SNPs contributed to a risk score. We studied the association of the risk score to lung function and COPD.**

***Measurements and Main Results*: Association with COPD was significant for three loci (*TNS1*, *GSTCD*, and *HTR4*) and the previously reported *HHIP* locus, and suggestive and directionally consistent for *AGER* and *TSHD4*. Compared with the baseline group (7 risk alleles), carrying 10–12 risk alleles was associated with a reduction in FEV_1 ($\beta = -72.21$ ml, $P = 3.90 \times 10^{-4}$) and FEV_1/FVC ($\beta = -1.53\%$, $P = 6.35 \times 10^{-6}$), and with COPD (odds ratio = 1.63, $P = 1.46 \times 10^{-5}$).**

***Conclusions*: Variants in *TNS1*, *GSTCD*, and *HTR4* are associated with COPD. Our highest risk score category was associated with a 1.6-fold higher COPD risk than the population average score.**

Keywords: FEV_1; FVC; genome-wide association study; modeling risk

Chronic obstructive pulmonary disease (COPD), characterized by airflow limitation that is not fully reversible, affects approximately 210 million people worldwide (1) and is among the leading causes of death in developed and developing countries (2, 3). Tobacco smoking is a potent cause of COPD, but not all smokers develop COPD and genetic determinants are also important (4). Identification of the genetic determinants of COPD could provide insight into molecular pathways that may be amenable to improved preventive and treatment strategies. A further potential utility of newly identified genetic associations is to predict disease risk. Current evidence available from common complex diseases where family history may be used, such as type 2 diabetes, suggests that tens of genetic variants with individually modest effects may provide similar but not necessarily substantially improved disease risk prediction over existing scores that incorporate family history (5, 6)

Genetic variants in *SERPINA1* that cause α_1-antitrypsin deficiency have long been known to affect COPD risk. Although there had been limited success in identifying additional susceptibility loci for COPD until 2009 (7), more recent genome-wide association (GWA) studies have shown associations between genetic loci and lung function measures that underpin the diagnosis of COPD (8–10). Two of these studies have investigated genome-wide association with lung function in large sample sizes (>20,000 subjects), focusing exclusively on quantitative lung function measures (8, 9). Since the advent of dense GWA genotyping platforms, the only loci convincingly associated with COPD have been *HHIP* (10, 11), which to date remains the strongest signal, *CHRNA3/5* (11), and *FAM13A* (12).

We hypothesized that genetic variants associated with FEV_1 and FEV_1/FVC would be associated with COPD. In a study of 20,288 individuals with GWA data and follow-up of top signals in a further 54,276 individuals (SpiroMeta Consortium), we previously identified five novel loci showing association ($P < 5 \times 10^{-8}$) with FEV_1 or FEV_1/FVC: in *TNS1* at 2q35, in *GSTCD* at 4q24, in *HTR4* at 5q33, in *AGER* at 6p21, and in *THSD4* at 15q23 (9).

In this article, based on the five loci reported by the SpiroMeta Consortium (9), we further investigate the clinical relevance of these loci. First, we test association of the sentinel single-nucleotide polymorphism (SNP) at each locus with COPD. Second, we investigate the combined effect of the risk alleles at all five novel loci described previously (*TNS1*, *GSTCD*, *HTR4*, *AGER*, and *THSD4*) and the previous association at 4q31 (near *HHIP*) (10) on lung function and COPD risk.

Some of the results of these studies have been previously reported in the form of abstracts (13, 14).

AT A GLANCE COMMENTARY

Scientific Knowledge on the Subject

Genome-wide association studies have reported novel loci for lung function, but the association of these loci with chronic obstructive pulmonary disease (COPD) has not yet been tested, and their effects in combination have not yet been documented.

What This Study Adds to the Field

We show associations between COPD and polymorphisms in *HTR4*, *GSTCD*, and *TNS1*. Using a six-SNP (single-nucleotide polymorphism) risk score, incorporating variants in *HTR4*, *GSTCD*, *TNS1*, *AGER*, *THSD4*, and near *HHIP*, the highest risk category (~5% of Europeans) is associated with a 1.6- fold risk of COPD compared with a common baseline group.

METHODS

Study 1: Single SNP Analysis of *TNS1*, *GSTCD*, *HTR4*, *AGER*, *THSD4* with COPD

Populations, phenotyping, and genotyping. Figure 1 shows the study populations and loci included in the study design. The study population consisted of 31,422 individuals over the age of 40 years from 12 population-based studies. These studies included the European Prospective Investigation into Cancer and Nutrition obese cases cohort (EPIC-obese case subjects) and population cohort (EPIC population-based), Generation Scotland: Scottish Family Health Study (GS: SFHS), Cooperative Health Research in the Region of Augsburg (KORA F4), Adult-onset Asthma and Nitric Oxide (ADONIX) Study, Busselton Health Study (BHS), British Regional Heart Study (BRHS), British Women's Heart and Health Study (BWHHS), Gedling Study (Gedling), Hertfordshire Cohort Study (HCS), Finnish Health 2000 Survey (Health 2000), Nottingham Smokers Study (Nottingham Smokers), and Medical Research Council National Survey of Health and Development (NSHD, or British 1946 Birth Cohort).

FEV_1 and the ratio of FEV_1 to FVC were measured in each study, using the spirometry methods detailed in the online supplement. The percent predicted FEV_1 was calculated according to previously described prediction equations (15, 16). Individuals with percent predicted FEV_1 less than 80% and FEV_1/FVC less than 0.7 (Global Initiative for Chronic Obstructive Lung Disease [GOLD] stages 2–4) were classified as COPD case subjects (17). Individuals with FEV_1 greater than 80% predicted and FEV_1/FVC greater than 0.7 were classified as control subjects. To minimize potential misclassification of COPD case subjects and control subjects, individuals not falling in either category (GOLD stage 1) were excluded from the analyses of COPD risk. We also excluded related individuals from the case–control analyses.

Genotyping was undertaken for a sentinel SNP at each of the following loci: *TNS1* (rs2571445), *GSTCD* (rs10516526), *HTR4* (rs3995090),

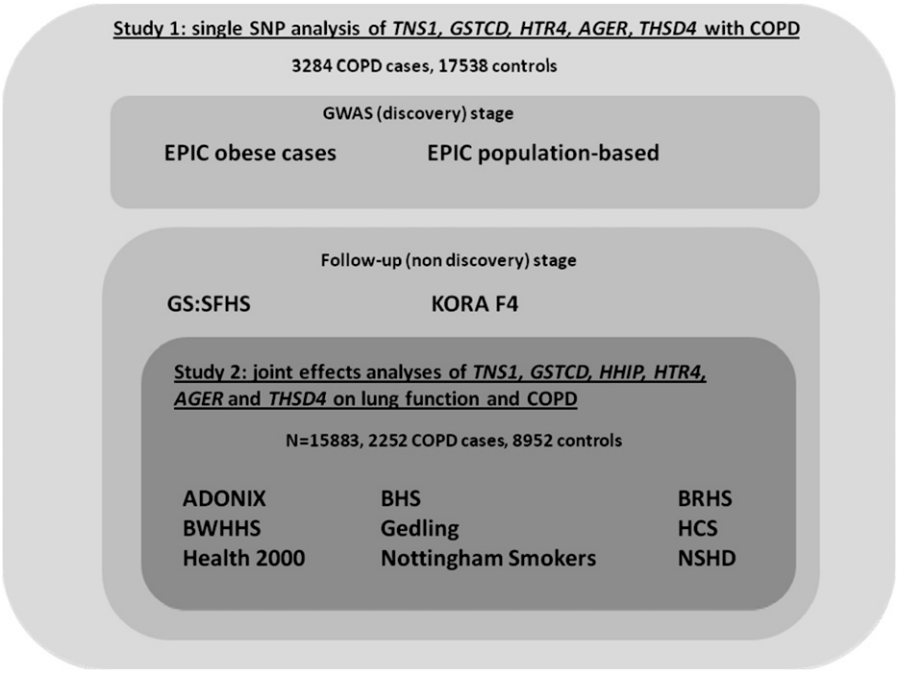

Figure 1. Study design. Single-nucleotide polymorphisms (SNPs) genotyped at each of the loci listed were rs2571445 (*TNS1*), rs10516526 (*GSTCD*), rs12504628 (near *HHIP*), rs3995090 (*HTR4*), rs2070600 (*AGER*), and rs12899618 (*THSD4*). Study 1 included all participating studies. Study 2 included the subset of studies that were genotyped for all six SNPs and that were not included in the discovery set of genome-wide association study (GWAS) data that led to the discovery of our five loci (*9*). Individuals excluded from study 2 chronic obstructive pulmonary disease (COPD) analyses were as follows: (*1*) individuals under the age of 40 years, (*2*) individuals with stage 1 COPD ($FEV_1/FVC < 0.7$ and percent predicted $FEV_1 > 80\%$), and (*3*) individuals with FEV_1/FVC greater than 0.7 but percent predicted FEV_1 less than 80%. Study abbreviations are as follows: EPIC obese case subjects (European Prospective Investigation into Cancer and Nutrition-obese case subjects) and EPIC population-based (European Prospective Investigation into Cancer and Nutrition cohort), GS:SFHS (Generation Scotland: Scottish Family Health Study), KORA F4 (Cooperative Health Research in the Region of Augsburg), ADONIX (Adult-onset Asthma and Nitric Oxide), BHS (Busselton Health Study), BRHS (British Regional Heart Study), BWHHS (British Women's Heart and Health Study), Gedling (Gedling Study), HCS (Hertfordshire Cohort Study), Health 2000 (Finnish Health 2000 Survey), Nottingham Smokers (Nottingham Smokers Study), and NSHD (Medical Research Council National Survey of Health and Development, also known as the British 1946 Birth Cohort).

AGER (rs2070600), and *THSD4* (rs12899618). Standard quality control approaches were used. To incorporate the previously reported locus near *HHIP* into the analyses, rs12504628 was genotyped in studies that employed KASPar genotyping (KBioscience, Hoddesdon, Herts, UK) and available *in silico* data for rs12504628 were used for EPIC, Health 2000, and a subset of BHS (footnote † in Table 1).

Statistical analysis. Each SNP genotype was coded 0, 1, or 2, corresponding to the number of copies of the coded allele. The effect estimates were oriented to the forward strand of the National Center for Biotechnology Information (NCBI) build 36 reference sequence of the human genome, using the alphabetically higher allele as the coded allele. For each of the five SNPs reported by the SpiroMeta Consortium (9), logistic regression was used within each study population to test association of the SNP with COPD. Adjustments for additional covariates were not used for this analysis given that the percent predicted FEV_1 (used to define COPD) includes adjustments for age, sex, and height (18). We defined Bonferroni-corrected statistical significance as P less than 0.01, to account for the testing of the five independent SNPs. After quality checks of the study population level data, pooled effect size estimates and their standard errors were computed across all studies, using an inverse variance weighting. Although previously shown to be associated with COPD and therefore not the focus of our analysis, for completeness we present the associations with rs12504628 (near *HHIP*).

Study 2: Joint Effects of *TNS1*, *GSTCD*, *HHIP*, *HTR4*, *AGER*, and *THSD4* Loci on Lung Function and COPD

Populations. The studies and loci included in study 2 are shown in Figure 1. We included the subset of studies that were genotyped for all six SNPs and that were not in the set of GWA data that led to the discovery of these loci (9). This restriction was to avoid cumulative biases resulting from individual SNP estimates of association that tend to be biased away from the null in "hits" from genome-wide discovery sets (winner's curse bias) (19). The phenotyping and genotyping of these studies were undertaken as described for study 1.

Statistical analysis: derivation of the risk score. To derive an unweighted risk allele score theoretically ranging from 0 to 12, each allele previously associated with reduced FEV_1 or FEV_1/FVC (9) contributed one to the risk score. The risk alleles for the six loci were as follows: A for rs2571445 (*TNS1*), A for rs10516526 (*GSTCD*), T for rs12504628 (*HHIP*), A for rs3995090 (*HTR4*), C for rs2070600 (*AGER*), and A for rs12899618 (*THSD4*). We categorized the risk scores into five groups; 0–4, 5–6, 7 (median number of risk alleles, baseline group), 8–9, and 10–12 risk alleles.

Statistical analysis: testing association between the risk score and lung function. We performed linear regressions of FEV_1 and FEV_1/FVC onto age, age squared, sex, and height to obtain residual phenotypes. Linear regressions were undertaken with each residual phenotype as the outcome variable, and an intercept and four indicator variables (for the four nonbaseline risk allele groups) as the explanatory variables to test for association. After quality checks of study level data, we computed pooled estimates using an inverse variance weighting.

Statistical analysis: testing association between the risk score and COPD. Individuals were classified as COPD case subjects (GOLD stages 2–4) or control subjects on the basis of the criteria described previously for study 1. To test the association of the unweighted risk allele score with COPD we used logistic regression with COPD as the outcome, an intercept, and the four indicator variables. Study level estimates were pooled after quality checks using an inverse variance weighting.

RESULTS

Study 1: Single SNP Analysis of *TNS1*, *GSTCD*, *HTR4*, *AGER*, and *THSD4* with COPD

The characteristics of the study participants are shown in Table 1. For the individual SNP associations with COPD a total of 3,284 individuals were classified as COPD case subjects (percent predicted $FEV_1 < 80\%$ and $FEV_1/FVC < 0.7$) and 17,538 individuals as control subjects ($FEV_1 > 80\%$ predicted and $FEV_1/FVC > 0.7$). Of the variants at the five loci shown to be associated with lung function (9), variants at three loci showed significant association with COPD: rs2571445 in *TNS1* (odds ratio [OR] per A allele, 1.10; 95% confidence interval [CI], 1.03–1.16; $P = 1.89 \times 10^{-3}$) (Figures 2 and 3), rs10516526 in *GSTCD* (OR per A allele, 1.24; 95% CI, 1.10–1.40; $P = 3.75 \times 10^{-4}$) (Figures 2 and 3), and rs3995090 in *HTR4* (OR per A allele, 1.12; 95%

TABLE 1. STUDY CHARACTERISTICS

Study Name	Subset	No. in Study 1	No. in Study 2	Male (*n*); Female (*n*)	Age: Mean (SD) (*yr*)	FEV_1: Mean (SD) (*L*)	FEV_1 Predicted: Mean (SD) (*L*)	FVC: Mean (SD) (*L*)	FEV_1/FVC Ratio: Mean (SD)	Never-smokers (n); Ever-smokers (n)	COPD Severity (% of COPD Case Subjects in GOLD Stage 3 or 4)	Genotyping
Study 1 Single SNP Analysis: GWAS (Discovery) Stage												
EPIC obese case subjects	All	1,104		476; 628	59.1 (8.8)	2.35 (0.69)	2.91 (0.62)	2.84 (0.87)	0.82(0.17)	489; 615		Affymetrix 500K
	Case subjects	75		47; 28	60.64 (8.45)	1.82 (0.67)	3.09 (0.63)	3.01 (1.05)	0.61 (0.09)	22; 53	30.14	
	Control subjects	599		252; 347	58.30 (8.76)	2.67 (0.62)	2.88 (0.61)	3.13 (0.79)	0.86 (0.07)	281; 318		
EPIC population-based	All	2,336		1,100; 1,236	59.2 (9.0)	2.50 (0.72)	2.95 (0.62)	3.04 (0.90)	0.85 (0.16)	1,061; 1,275		Affymetrix 500K
	Case subjects	190		105; 85	62.31 (8.54)	1.81 (0.62)	2.95 (0.63)	3.00 (0.96)	0.60 (0.09)	72; 118	20.11	
	Control subjects	1,442		677; 765	58.76 (8.81)	2.78 (0.64)	2.94 (0.62)	3.29 (0.82)	0.85 (0.08)	709; 733		
Study 1 Single SNP Analysis: Follow-up (Nondiscovery) Stage												
GS:SFHS	All	5,474		2,254; 3,220	46.0 (14.3)	3.15 (0.87)	3.32 (0.75)	4.11 (1.03)	0.77 (0.10)	3,005; 2,469		TaqMan
	Case subjects	335		118; 217	58.4 (9.2)	1.89 (0.54)	2.89 (0.59)	3.32 (0.85)	0.58 (0.10)	123; 212	11.94	
	Control subjects	2,567		1,053:1,514	53.2 (8.5)	3.12 (0.72)	3.11 (0.63)	3.99 (0.91)	0.78 (0.07)	1,457; 1,110		
KORA F4	All	1,305		610; 695	51.6 (5.7)	3.32 (0.81)	3.29 (0.63)	4.28 (1.00)	0.78 (0.06)	499; 806		TaqMan
	Case subjects	59		30; 29	53.4 (6.0)	2.14 (0.66)	3.24 (0.64)	3.47 (0.96)	0.61 (0.06)	12; 47	10.17	
	Control subjects	1,109		512; 597	51.5 (5.7)	3.45 (0.76)	3.28 (0.62)	4.36 (0.96)	0.79 (0.04)	456; 653		
Study 1 Single SNP Analysis and Study 2 Joint Effects Analyses: Follow-up (Nondiscovery) Stage												
ADONIX	All	1,423	1,282	669; 754	49.1 (13.5)	3.34 (0.86)	3.23 (0.66)	4.24 (1.02)	0.79 (0.07)	798; 625		KASPar*
	Case subjects	46	41	27; 19	55.7 (9.3)	2.02 (0.57)	3.23 (0.66)	3.35 (0.87)	0.60 (0.07)	12; 34	13.04	
	Control subjects	783	711	361; 422	61.4 (8.4)	3.23 (0.73)	3.23 (0.67)	4.08 (0.91)	0.79 (0.04)	448; 335		
BHS	All†	4,350	787	1,793; 2,557	50.1 (17.0)	3.02 (0.97)	3.18 (0.82)	3.89 (1.16)	0.77 (0.08)	2,459; 1,891		TaqMan
	Case subjects†	200	92	132; 68	66.9 (11.6)	1.60 (0.60)	2.85 (0.66)	2.73 (0.91)	0.58 (0.09)	67; 133	19.5	
	Control subjects†	2,307	386	944; 1,363	57.9 (12.3)	2.87 (0.83)	2.93 (0.73)	3.66 (1.05)	0.78 (0.05)	1,387; 920		
BRHS	All	3,877	3,415	3,877; 0	68.7 (5.5)	2.57 (0.69)	3.03 (0.4)	3.37 (0.84)	0.77 (0.12)	1,125; 2,752		KASPar*
	Case subjects	641	572	641; 0	69.7 (5.4)	1.76 (0.51)	3 (0.4)	3.01 (0.8)	0.59 (0.09)	111; 530	28.39	
	Control subjects	2,168	1,905	2,168; 0	68.3 (5.5)	2.96 (0.48)	3.03 (0.4)	3.65 (0.65)	0.82 (0.07)	760; 1,408		
BWHHS	All	3,644	3,319	0; 3,644	68.8 (5.5)	1.98 (0.52)	2.16 (0.31)	2.82 (0.76)	0.71 (0.09)	2,060; 1,584		KASPar*
	Case subjects	659	600	0; 659	69.8 (5.4)	1.36 (0.35)	2.14 (0.3)	2.32 (0.54)	0.59 (0.08)	253; 406	15.63	
	Control subjects	1,808	1,653	0; 1,808	68.2 (5.4)	2.23 (0.41)	2.18 (0.3)	2.93 (0.56)	0.76 (0.05)	1,153; 655		
Gedling	All	1,263	1,188	632; 631	56.2 (12.3)	2.85 (0.85)	3.07 (0.69)	3.68 (1.02)	0.77 (0.07)	633; 630		KASPar*
	Case subjects	103	98	67; 36	66.2 (9.1)	1.73 (0.61)	2.88 (0.66)	2.82 (0.83)	0.61 (0.09)	21; 82	24.27	
	Control subjects	840	789	417; 423	57.3 (9.8)	3 (0.73)	3.03 (0.65)	3.8 (0.9)	0.79 (0.04)	431; 409		
HCS	All	2,850	2,343	1,511; 1,339	66.1 (2.8)	2.44 (0.68)	2.80 (0.55)	3.42 (0.92)	0.72 (0.09)	1,319; 1,531		KASPar*
	Case subjects	536	441	308; 228	66.3 (2.8)	1.84 (0.55)	2.87 (0.56)	2.09 (0.85)	0.60 (0.09)	159; 377	15.1	
	Control subjects	1,519	1,264	758; 761	66.0 (2.9)	2.67 (0.60)	2.75 (0.54)	3.51 (0.82)	0.76 (0.04)	837; 682		
Health 2000	All	888	882	427; 456	50.2 (11.0)	3.32 (0.91)	3.14 (0.67)	4.19 (1.08)	0.79 (0.07))	266; 617		Illumina 610K
	Case subjects	32	32	20; 12	60.91 (8.83)	1.78 (0.68)	3.05 (0.66)	3.05 (0.95)	0.58 (0.10)	5; 27	37.5	
	Control subjects	580	580	256; 324	53.19 (8.31)	3.28 (0.77)	3.15 (0.67)	4.09 (0.97)	0.80 (0.05)	192; 388		

(Continued)

TABLE 1. (CONTINUED)

Study Name	Subset	No. in Study 1	No. in Study 2	Male (*n*); Female (*n*)	Age: Mean (SD) (*yr*)	FEV_1: Mean (SD) (*L*)	FEV_1 Predicted: Mean (SD) (*L*)	FVC: Mean (SD) (*L*)	FEV_1/FVC Ratio: Mean (SD)	Never-smokers (n); Ever-smokers (n)	COPD Severity (% of COPD Case Subjects in GOLD Stage 3 or 4)	Genotyping
Nottingham Smokers	All	509	466	280; 229	59.5 (10.4)	2.00 (0.95)	2.98 (0.61)	3.02 (1.06)	0.64 (0.16)	0; 509		KASPar*
	Case subjects	242	227	145; 97	63.2 (9.5)	1.28 (0.57)	2.87 (0.59)	2.5 (0.87)	0.51 (0.12)	0; 242	64.46	
	Control subjects	153	138	70; 83	54.8 (8.9)	2.89 (0.61)	3.08 (0.62)	3.69 (0.81)	0.79 (0.05)	0; 153		
NSHD	All	2,404	2,201	1,206; 1,198	53 (0)	2.80 (0.70)	3.20 (0.56)	3.51 (0.89)	0.80 (0.09)	1,003; 1,401		KASPar*
	Case subjects	166	149	102; 64	53 (0)	2.11 (0.58)	3.35 (0.54)	3.46 (0.89)	0.61 (0.08)	49; 117	15.06	
	Control subjects	1,663	1,526	848; 815	53 (0)	3.03 (0.62)	3.20 (0.56)	3.69 (0.81)	0.83 (0.06)	765; 898		
Total	All	31,422	15,883									
	Case subjects	3,284	2,252									
	Control subjects	17,538	8,952									

Definition of abbreviations: ADONIX = Adult-onset Asthma and Nitric Oxide Study; BHS = Busselton Health Study; BRHS = British Regional Heart Study; BWHHS = British Women's Heart and Health Study; COPD = chronic obstructive pulmonary disease; EPIC = European Prospective Investigation into Cancer and Nutrition; GOLD = Global Initiative for Chronic Obstructive Lung Disease; GS:SFHS = Generation Scotland: Scottish Family Health Study; GWAS = genome-wide association study; HCS = Hertfordshire Cohort Study; KORA F4 = Cooperative Health Research in the Region of Augsburg; NSHD = Medical Research Council National Survey of Health and Development.

* KASPar genotyping (KBiosciences, Hoddesdon, Herts, UK; http://www.kbioscience.co.uk/).

† The BHS had genotype data for *HHIP* only for a subset of individuals (n = 1,168, 131 COPD case subjects and 565 control subjects); this is therefore the subset included in study 2.

CI, 1.05–1.18; $P = 1.79 \times 10^{-4}$) (Figures 2 and 3). The associations with COPD were in the direction expected, given the reported direction of these allelic effects on lung function (Figure 2) (9). The point estimates of the effects on COPD for the sentinel SNPs in *AGER* (rs2070600) and in *THSD4* (rs12899618) were of the magnitude and direction expected, but did not reach statistical significance. Although it was not a primary aim of our study to investigate the association between COPD and the locus near *HHIP* at 4q31 (an association reported by several studies [10, 11, 20]), we were also able to test this in a subset of our data (2,890 case subjects and 13,862 control subjects) for which the rs12504628 genotype was available. We confirmed the association between the 4q31 locus and COPD: OR per rs12504628 T allele, 1.19 (95% CI, 1.12–1.27) and $P = 4.55 \times$

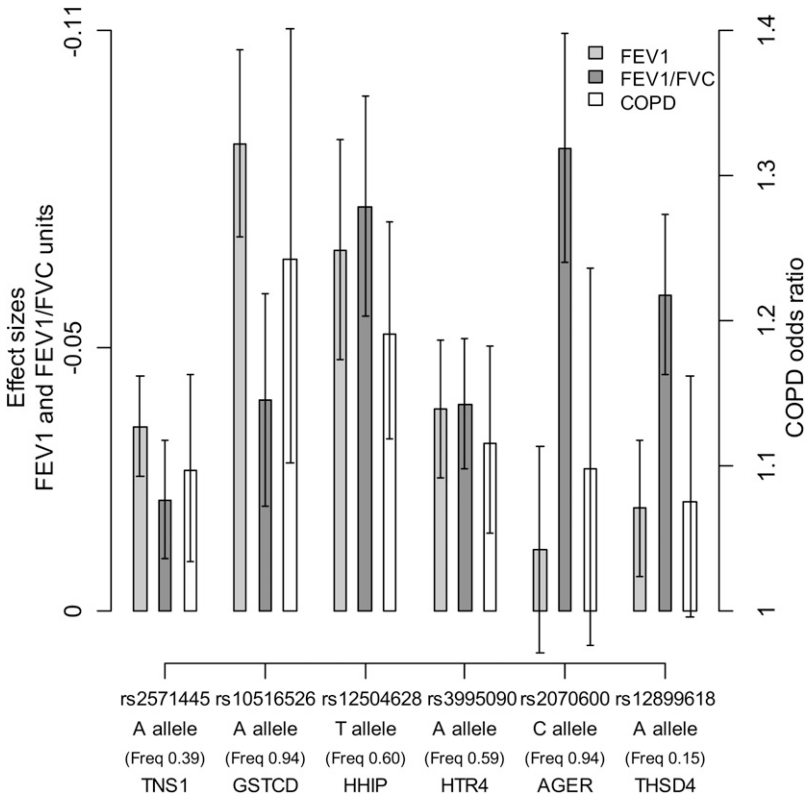

Figure 2. Association of sentinel single-nucleotide polymorphisms (SNPs) at five novel lung function loci and a previously reported *HHIP* SNP with chronic obstructive pulmonary disease (COPD), and comparison with reported associations with FEV_1 and FEV_1/FVC. Shown are the results for COPD after testing for association in 3,284 COPD case subjects and 17,538 control subjects, and a comparison with the associations with FEV_1 and FEV_1/FVC in the combined discovery and follow-up data reported by Repapi and colleagues (9). *Boxes* indicate the point estimates of the effect sizes and *whiskers* the 95% confidence intervals.

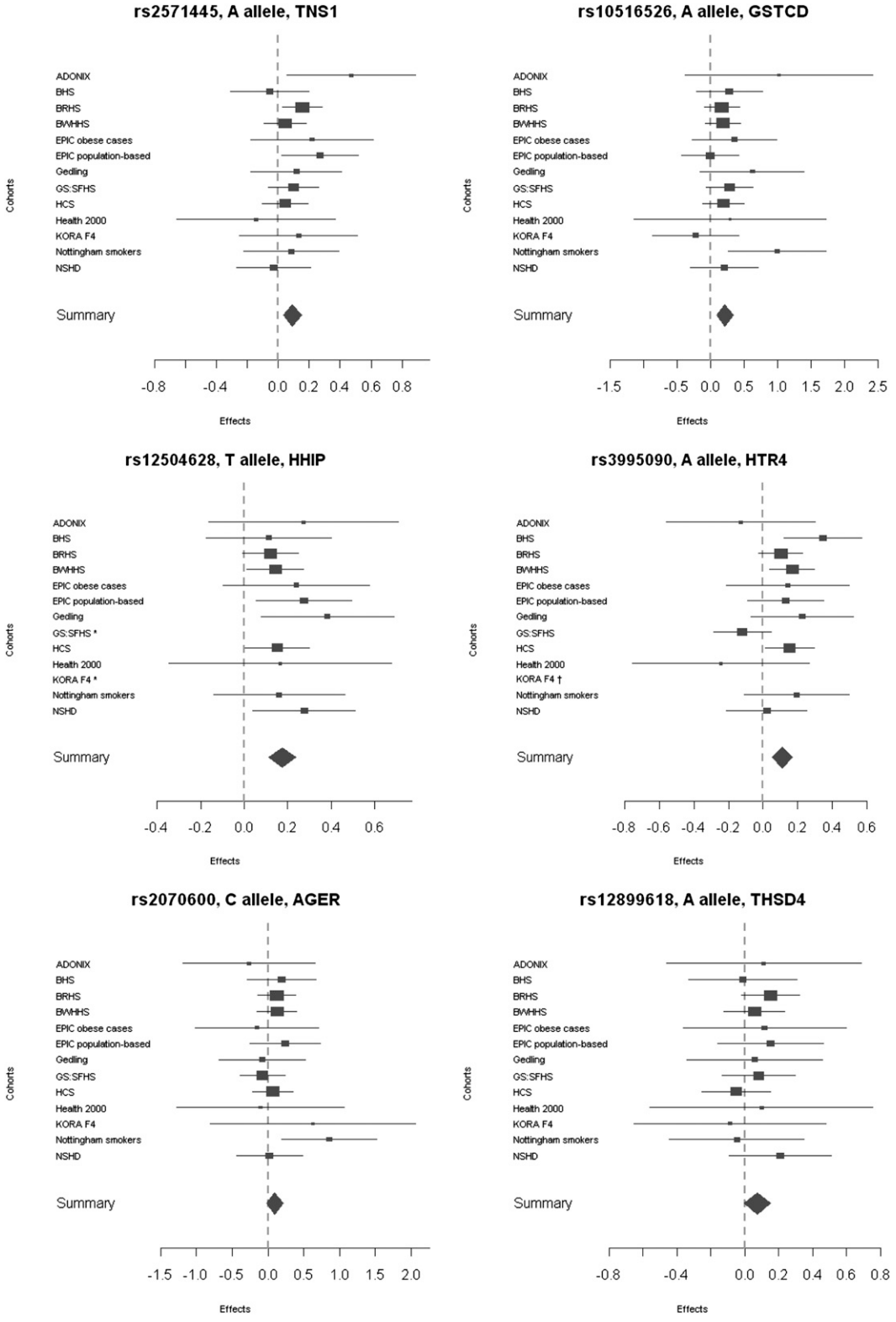

Figure 3. Forest plots of the meta-analysis of association tests with chronic obstructive pulmonary disease (COPD) for the six loci (*TNS1*, *GSTCD*, *HHIP*, *HTR4*, *AGER*, and *THSD4*). *rs12504628 (*HHIP*) data were not available in GS:SFHS or KORA F4. †KORA F4 failed to genotype rs3995090 (*HTR4*). ADONIX = Adult-onset Asthma and Nitric Oxide Study; BHS = Busselton Health Study; BRHS = British Regional Heart Study; BWHHS = British Women's Heart and Health Study; COPD = chronic obstructive pulmonary disease; EPIC = European Prospective Investigation into Cancer and Nutrition; GS:SFHS = Generation Scotland: Scottish Family Health Study; HCS = Hertfordshire Cohort Study; KORA F4 = Cooperative Health Research in the Region of Augsburg; NSHD = Medical Research Council National Survey of Health and Development.

10^{-8}. We tested for heterogeneity of effect sizes across the studies for each SNP; a chi-squared heterogeneity test was not statistically significant ($P > 0.10$) for any of the six SNPs.

As expected, a much higher proportion were ever-smokers among COPD case subjects (72%) than among control subjects (49%). This raises the possibility that these variants could influence COPD risk via an effect on smoking behavior. Therefore we explored whether the effects on COPD risk could be mediated via smoking. First, we examined the effects of an additional pack-years adjustment in smokers in a subset of our data with pack-years available and obtained similar results to those without adjustment for pack-years (*see* Figure E1 in the online supplement). Second, we assessed the association between these SNPs and two smoking behavior phenotypes in the Oxford-GlaxoSmithKline (Ox-GSK) consortium data set (21). As shown in Table E1, none of the SNPs showed even nominal

association ($P < 0.05$) in 18,598 ever-smokers versus 15,041 never-smokers and none of the SNPs showed nominal association with the number of cigarettes smoked per day (n = 15,574). This evidence strongly suggests that the effects of these variants on lung function and COPD risk are not mediated via tobacco addiction.

Study 2: Joint Effects of *HHIP*, *TNS1*, *GSTCD*, *HTR4*, *AGER*, and *THSD4* Loci on Lung Function and COPD

The characteristics of the study participants are shown in Table 1. In all, 15,883 individuals were included in the analyses of FEV_1 and FEV_1/FVC (Table 1). A trend in lung function (FEV_1 and FEV_1/FVC) was shown across risk allele categories (Figure 4). Notably, compared with the baseline group of seven risk alleles, carrying 10–12 risk alleles was associated with a reduction in FEV_1 (coefficient –72.21 ml [95% CI, –112.12 to –32.30 ml]; $P = 3.90 \times 10^{-4}$) and a reduction in FEV_1/FVC (coefficient –1.53% [95% CI, –2.20% to –0.87%]; $P = 6.35 \times 10^{-6}$). Approximately 5% of the study population carried 10–12 risk alleles and 28% of the study population carried seven risk alleles. The magnitude of decline in FEV_1 compared with the baseline group is equivalent to the physiological average ageing decline in lung function over approximately four years in a nonsmoking population (22).

To assess the joint effects of the risk alleles on COPD, data from nondiscovery cohorts (Figure 1) with genotype data available on all six SNPs were used. This included 2,252 COPD case subjects (percent predicted $FEV_1 < 80\%$ and $FEV_1/FVC < 0.7$) and 8,952 control subjects ($FEV_1 > 80\%$ predicted and $FEV_1/FVC > 0.7$) as shown in Figure 1 and Table 1. Again, a clear trend in COPD risk across the categories was observed (Figure 4). Compared with a baseline of seven risk alleles, carrying 10 to 12 risk alleles was associated with an increased risk of COPD (OR, 1.63; 95% CI, 1.31–2.03; $P = 1.46 \times 10^{-5}$) (Figure 4).

DISCUSSION

In a study of the set of novel variants we reported to be associated with lung function (9), we show significant association with COPD for three of the five loci—in *HTR4*, *GSTCD*, and *TNS1*—emphasizing the clinical relevance of these loci to respiratory disease. We also confirm the association between the 4q31 locus near *HHIP* and COPD, and show expected direction and magnitude of effect (although nonsignificant) for the two remaining of the five lung function loci studied. In addition, we provide an estimate of the combined effect sizes of these loci on lung function and COPD in studies independent of the data used to discover these associations. We show that the highest number of risk alleles (10–12 risk alleles, 5% of our population) is associated with a 1.6-fold elevation of COPD risk, compared with a common baseline group of individuals with 7 risk alleles (28% of our population).

The loci associated with COPD may provide important insights into the pathways underlying the development of COPD. The sentinel SNP at the 4q24 locus (rs10516526) is intronic in *GSTCD*, which encodes a glutathione *S*-transferase, C-terminal domain–containing protein. This protein may be involved in cellular detoxification, catalyzing conjugation of glutathione to products of oxidative stress (23), and in regulating the synthesis of prostaglandins and leukotrienes (23). *GSTCD* also shows homology with chloride intracellular channels and could therefore influence lung function via other molecular pathways (9). A second locus associated with COPD was localized to *TNS1* at 2q35, where the sentinel variant was a nonsynonymous coding SNP. Tensin-1 is an actin-binding protein with SH2 (Src homology-2) domains, possibly involved in signal transduction (24) and cell migration (25). At 5q33, the sentinel variant was an intronic SNP (rs3995090) in *HTR4* encoding the 5-hydroxytryptamine receptor-4 (HTR4), which is expressed in neurons and airway epithelial type II cells, where it may

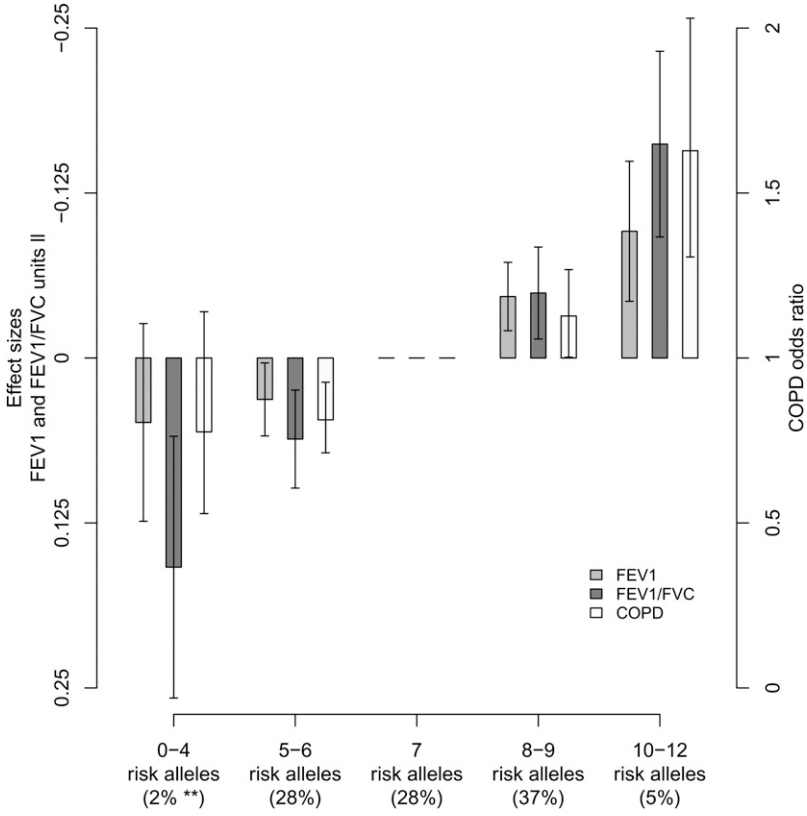

Figure 4. Association of risk scores with lung function and chronic obstructive pulmonary disease (COPD). The risk score theoretically ranges from 0 to 12 across the six loci (*TNS1*, *GSTCD*, *HHIP*, *HTR4*, *AGER*, and *THSD4*). The risk allele category was in each case compared with a baseline of seven risk alleles. *Boxes* indicate the point estimates of the effect sizes and *whiskers* the 95% confidence intervals. ‖To facilitate the plotting of the effect size estimates for FEV_1 and FEV_1/FVC on the same axes, effect sizes are given in terms of the proportion of a standard deviation of FEV_1 and FEV_1/FVC; we used a standard deviation of 752 ml for FEV_1 and 9.45% for FEV_1/FVC (obtained as weighted averages across studies). **Proportion of individuals within each risk score category.

regulate cytokine responses (26). In contrast, the sentinel SNP at 4q31 reported in our study (rs1250468) and that reported in previous studies (rs13147758, $r^2 = 0.97$ with rs1250468) (10, 11), lies in an intergenic region upstream of *HHIP*. The hedgehog (Hh) gene family encodes signaling molecules involved in regulating lung morphogenesis (27), although this locus has been previously associated with height (28), implicating a role in skeletal growth and development.

Although we did not demonstrate a statistically significant association between COPD and two loci, the estimated effects were in the expected direction. These included a nonsynonymous coding SNP (rs2070600) in *AGER*, within a gene-rich region of the major histocompatibility complex (6q21), and an intronic SNP (rs12899618) in *TSHD4* (15q23). *AGER* is a strong candidate as it is highly expressed in the lung (29) and altered *AGER* expression has been noted in COPD lung tissue (30) and in subjects with idiopathic pulmonary fibrosis (31). Similarly the gene product of *THSD4* shows homology with the thrombospondin family of extracellular calcium-binding proteins implicated in wound healing, inflammation, and angiogenesis (32) and could also be a good candidate for COPD development. Of the five new loci (9) we examined, the loci showing association with COPD in our data showed strongest association with FEV_1 (*HTR4*, *GSTCD*, and *TNS1*), whereas the two loci that did not show a significant association with COPD (*AGER* and *THSD4*) showed association with FEV_1/FVC, but without a strong association with FEV_1. A simpler explanation for these different findings may be limited statistical power to detect the modest effects of these common genetic variants on a binary outcome.

The effect sizes of the SNPs at the three loci associated with COPD (*HTR4*, *GSTCD*, and *TNS1*) and the locus previously reported (*HHIP*) are modest, in the range 1.10–1.24 per copy of the risk allele at each locus. For the previously reported 4q31 locus near *HHIP*, we estimate an odds ratio of 1.19 (95% CI, 1.12–1.27) for COPD. Although larger effect sizes have been described, for example, an odds ratio for COPD equivalent to approximately 1.4 for the risk allele A at rs13118928 in the U.S. National Emphysema Treatment Trial/Normative Ageing Study and Bergen populations (11); the Rotterdam Study (20) and Framingham Heart Study (10) described odds ratios of 1.25 and 1.10 for this allele. Although the differences in the effect size estimates for the 4q31 locus could be attributable to differences in study characteristics, such as age, these differences could be explained in part by winner's curse bias (19), in which the effect size is generally overestimated in the study that first detects the association at the stringent levels of significance required in GWA studies. Additional evidence from independent populations, such as that provided from our study, can be important in establishing effect size estimates that are likely to be unaffected by such biases.

Most of the constituent studies did not measure postbronchodilation spirometry, and therefore our analysis was limited to prebronchodilation spirometry measures. In the Nottingham Smokers Study, in which both pre- and postbronchodilation spirometry were measured, the positive predictive value of prebronchodilation-defined COPD for diagnosis of postbroncholidation-defined COPD was 98% (Table E2). In all association tests reported in this article, we excluded individuals with GOLD stage 1 COPD from both case subjects and control subjects. As has been shown previously (33), the use of prebroncholidation spirometry to diagnose COPD when GOLD stage 1 COPD case subjects are included would lead to substantial misclassification (e.g., in the Nottingham Smokers Study the positive predictive value would decline to 89%; Table E3). Demonstration of airflow obstruction with postbronchodilator spirometry is required to make a formal diagnosis of COPD. Patients with partly or fully reversible airflow obstruction may have fundamentally different pathological processes contributing to airflow obstruction. One potential bias is that there may be some cases with misclassification of either an asthma diagnosis or the COPD diagnosis. In addition, both asthma and COPD are common diagnoses and may coexist in the same individuals. The associations with lung function that we show were found in the general population and if the SNP effects are specific to asthma and not present in COPD, then the inclusion of patients with asthma could overestimate the contribution of the SNPs to COPD. However, we examined the effects of exclusion of patients with known asthma from the case subjects in a subset of the data with asthma diagnosis available and this did not alter the findings substantially (Figure E2). In addition, within the Nottingham Smokers Study, for which we have in-depth phenotype data and know that individuals with asthma and never-smokers with airflow obstruction have been specifically excluded, we found similar effect estimates. The effect size estimates in Nottingham Smokers were also consistent with effects on COPD for our sentinel SNPs in *HTR4*, *GSTCD*, and *TNS1* (Figure 3).

Various criteria have been proposed for the classification of COPD. Our main analysis was based on GOLD criteria (17). As misclassification is particularly likely to occur when there is poor separation of criteria for case subjects and control subjects, we excluded from our analyses altogether any individuals with mild COPD (stage 1 GOLD criteria: $FEV_1/FVC < 70\%$ but $FEV_1 \geq 80\%$ predicted).

Our conclusions were not materially altered by classifying cases instead on the basis of lower limit of normal equations (18) (Figure E3). Misclassification could also occur as a consequence of spirometry measurement error, and this too would tend to lead to underestimation of the effects of the SNPs studied. An important limitation of our study is that the misclassification and smoking exposure cannot be fully addressed, using these cohorts, because postbronchodilator spirometry and quantification of tobacco smoke exposure is not captured in several of our cohorts. Further studies of patients with COPD with more detailed smoking exposure measurements and more extensive respiratory phenotyping should provide further insight into the precise effects of these SNPs on COPD risk and COPD progression.

We focused on the study of SNPs in six loci: *TNS1*, *GSTCD*, *HHIP*, *HTR4*, *AGER*, and *THSD4.* Further studies will be required to investigate the potential association with COPD of SNPs in the regions of *GPR126*, *ADAM19*, *PTCH*, and *PID1* (additional regions described by the Cohorts for Heart and Aging Research in Genomic Epidemiology [CHARGE] Consortium) (8). The addition of these SNPs, the sentinel SNP from *FAM13A* (associated with COPD and lung function [8, 12]), *CHRNA3/5* (11), and other new loci from ongoing powerful genome-wide association studies to the risk score we constructed would be expected to improve the discrimination of such a score for prediction of COPD. With some exceptions, such as the major histocompatibility region, genome-wide association studies have been successful in localizing association signals (34, 35), but further research is likely to lead to improved resolution of association signals and possibly to detection of multiple independent causal variants at each of these loci. The simple risk score we present here would need to be adapted to incorporate an appropriate weighting as the score incorporates more SNPs, with a greater range of effect sizes. However, for the limited number of SNPs we examined, the inclusion of a weighting (based on the allelic effect size estimate) did not materially alter our findings. We also constructed risk scores in a subset of the data, measuring the effect on COPD *per risk allele* in all individuals and in ever-smokers only; the odds ratios were 1.14 and 1.13, respectively, suggesting that

the effect of these loci on COPD is similar in the general population and in ever-smokers only. As more complete risk scores are constructed, it will be important to investigate their predictive potential in subgroups stratified by smoking status, and to examine whether such a score can improve on risk prediction from conventional risk factors alone (including age, sex, smoking status, history of asthma, and family history, where available).

Our study demonstrates that some loci underlying lung function are associated with COPD, and provides a proof of concept that investigation of genetic determinants of lung function can be a successful strategy to discover molecular pathways underlying COPD. Pathways involving *HTR4*, *GSTCD*, and *TNS1* are strong candidates for potential interventions to prevent or alleviate COPD. We show, in studies unaffected by winner's curse bias, that the highest risk category of a six-SNP risk score is associated with a 1.6-fold risk of COPD compared with a common baseline group of seven risk alleles. It is not yet known whether these SNPs could contribute to a clinically useful strategy for prediction and intervention to prevent COPD, although the high absolute risk of COPD in smokers would suggest that the clinical and public health impact of such an approach is worthy of investigation as more loci are discovered and incorporated into risk scores.

Author Disclosure: I.S. received a sponsored grant from Pfizer. I.P.H. is a board member of the MRC Clinical Training Panel and has received lecture fees from Pfizer. M.A.S., L.W., E.R., M.O., P.B., T.J., J.Z., E.A., A.D., S.K., B.S., G.C., J.H., L.J.P., A.H., S.W., P.W., S.E., G.S., I.B., R.L., N.W., C.C., E.D., S.S., J.L., J.M., S.D., Å.N., A. C.O., S.K., J.H., H.S., T.M.M., I.P., M.H., S.R., I.S., J.B., M.K., J.B., F.N., J.H., D.L., R. M., A.L.J., C.M.J., and M.T. do not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript.

Acknowledgment: The SpiroMeta Consortium acknowledges ENGAGE (Engaging ICT cooperation between Europe and the Asia-Pacific region).

Author Contributions

ADONIX: *investigators*—Åsa Torinsson Naluai, Fredrik Nyberg, Anna-Carin Olin, Santosh Dahgam; *contributions*—project conception, design, and management: A.-C.O.; phenotype collection and data management: A.-C.O.; genotyping: A.T. N., F.N.; data analysis: A.T.N., F.N., A.-C.O, S.D. **BHS:** *investigators*—Gemma Cadby, Jennie Hui, Alan L. James, Lyle J. Palmer; *contributions*—project conception, design, and management: A.L.J., L.J.P.; phenotype collection and data management: G.C., A.L.J., L.J.P.; data analysis: G.C., J. Hui, L.J.P.; **BRHS:** *investigators*—Aroon D. Hingorani, Richard W. Morris, S. Goya Wannamethee, Peter H. Whincup; *contributions*—phenotype collection and data management: R.W.M., S.G.W., P.H.W.; genotyping: A.D.H., R.W.M., P.H.W.; data analysis: R.W.M. **BWHHS:** *investigators*—George Davey Smith, Shah Ebrahim, Debbie A. Lawlor, Peter H. Whincup; *contributions*—phenotype collection and data management: G.D.S., S. E., D.A.L., P.H.W.; data analysis: D.A.L. **EPIC:** *investigators*—Inês Barroso, Ruth J. F. Loos, Nicholas J. Wareham, Jing Hua Zhao. *contributions*—project conception, design, and management: I.B., R.J.F.L., N.J.W., J.H.Z.; phenotype collection and data management: N.J.W.; genotyping: I.B., R.J.F.L., N.J.W., J.H.Z.; data analysis: R.J.F.L., J.H.Z. **Gedling:** *investigators*—John R. Britton, Tricia M. McKeever, Ian D. Pavord; *contributions*—phenotype collection and data management: J.R.B., T.M.M., I.D.P.; data analysis: M.O., M.D.T. **GS:SFHS:** *investigators*—Cathy Jackson, Shona Kerr, Anna Dominiczak, Blair Smith; *contributions*—project conception, design, and management: C.J., A.D., S.K., B.S.; phenotype collection and data management: C.J., B. S., A.D.; genotyping: S.K. **HCS:** *investigators*—Cyrus Cooper, Elaine Dennison, John W. Holloway, Seif Shaheen; *contributions*—project conception, design, and management: C.C., E.D., J.W.H.; phenotype collection and data management: C.C., E.D., S.S.; genotyping: C.C., E.D., J.W.H.; data analysis: J.W.H., S.S. **Health 2000:** *investigators*—Markku Heliövaara, Mika Kähönen, Samuli Ripatti, Ida Surakka; *contributions*—project conception, design, and management: M.H., M.K.; phenotype collection and data management: M.H., M.K.; genotyping: S. R., I. Surakka; data analysis: M.K., S.R., I. Surakka. **KORA F4:** *investigators*—Eva Albrecht, Stefan Karrasch, Joachim Heinrich, Holger Schulz; *contributions*—project conception, design, and management: S.K., J. Heinrich, H.S.; phenotype collection and data management: S.K., J. Heinrich, H.S.; data analysis: E.A. **Nottingham Smokers:** *investigators*—John D. Blakey, Ian P. Hall, Ma'en Obeidat, Ian Sayers; *contributions*—project conception, design, and management: I.P.H., M.O., I. Sayers; phenotype collection and data management: J.D.B., I.P.H., M.O., I. Sayers; genotyping: I.P.H.; data analysis: M.O., I. Sayers, M.D.T. **NSHD:** *investigators*—The "NSHD Respiratory Study Team" team members involved were as follows: Zaina Al-Kanaani, Anna Hansell, Rebecca Hardy, Diana Kuh, Andrew Wong; *contributions*—phenotype collection and data management: R.H., D.K., A.W.; genotyping: D.K., A.W.; data analysis: Z.A.-K., A.H., R.H., A.W. **Ox-GSK:** *investigators*—Jason Z. Liu, Jonathan Marchini; *contributions*—project conception, design, and management: J.M.; data analysis: J.Z.L.***Manuscript Conception, Design, and Management***: I.P.H., M.S.A., M.D.T.

Data Analysis and Bioinformatics: P.R.B., I.P.H., T.J., M.O., E.R. I.S., M.S.A., M.D.T., L.V.W.***Writing the Manuscript***: I.P.H., M.S.A., M.D.T.***Cohort Funding*****ADONIX:** The ADONIX Study was funded by the Swedish Research Council for Worklife and Social Research (FAS), grants 2001-0263, 2003-0139, Swedish Heart and Lung Foundation grant 20050561 and a collaborative research grant from AstraZeneca. **BHS:** The Busselton Health Study acknowledges the generous support for the 1994/5 follow-up study from Healthway, Western Australia. The Busselton Health Study is supported by The Great Wine Estates of the Margaret River region of Western Australia. The BHS gratefully acknowledges the assistance of the Western Australian DNA Bank (NHMRC Enabling Facility) with DNA samples and the support provided by the Western Australian Genetic Epidemiology Resource (NHMRC Enabling Facility) for this study. **BRHS:** Blood collection and spirometric measures were performed in 1998–2000 as part of a field study funded by the British Heart Foundation (grant PG/97012). DNA was extracted by Geneservice Limited. A.D.H. has a British Heart Foundation Senior Research Fellowship (FS05/125). **BWHHS:** The British Women's Heart and Health Study (BWHHS) is funded by the U.K. Department of Health Policy Research Program and the British Heart Foundation. A separate British Heart Foundation project grant (PG/06/154/22043) funded cotinine assays. **EPIC:** The EPIC Norfolk Study is funded by Cancer Research UK and the Medical Research Council. I.B. acknowledges funding from the Wellcome Trust (077016/Z/05/Z) and from the U.K. NIHR Cambridge Biomedical Research Centre. **Gedling:** The Nottingham Gedling cohort collection was funded by Asthma UK and the British Lung Foundation. **GS:SFHS:** The Generation Scotland: Scottish Family Health Study is funded by the Chief Scientist Office, part of the Scottish Government Health Directorate (http://www.sehd.scot.nhs.uk/cso), grant reference number CZD/16/6. Genotyping was conducted by the Genetics Core at the Wellcome Trust Clinical Research Facility, University of Edinburgh, Western General Hospital, Edinburgh, UK. **HCS:** The Hertfordshire Cohort Study DNA collection was funded by the Medical Research Council, Arthritis Research Campaign, and International Osteoporosis Foundation. **Health 2000:** This study was financially supported by the Medical Research Fund of the Tampere University Hospital. **KORA F4:** The KORA Augsburg studies were financed by the Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany and supported by grants from the German Federal Ministry of Education and Research (BMBF) in the context of the German National Genome Research Network (NGFN-2 and NGFN-plus). Our research was supported within the Munich Center of Health Sciences (MC Health) as part of LMUinnovativ. **Nottingham Smokers:** The Nottingham Smokers cohort collection was funded by the University of Nottingham. **NSHD:** The Medical Research Council National Survey of Health and Development (NSHD, also known as the British 1946 Birth Cohort) is funded by the Medical Research Council. The MRC Unit for Lifelong Health and Ageing is responsible for NSHD. Z.A.-K., R.H., D.K., and A.W. are all affiliated with the MRC National Survey of Health and Development, MRC Unit for Lifelong Health and Ageing, London, UK except A.H., who is affiliated with the MRC-HPA Centre for Environment and Health, Imperial College London. Z.A.-K. has a studentship funded by Department of Health Air Pollution PRP Grant Ref. No. 0020029. **Ox-GSK:** GlaxoSmithKline (GSK), a pharmaceuticals company that is interested in developing therapies for lung disease and new cessation therapies for smoking, funded a postdoctoral fellowship for J.Z.L. at Oxford University. GSK also funded the collection, characterization, and, in some cases, the genotyping and genotype data preparation for several of the cohorts used in this study. A. Roses and P. Matthews played crucial roles in establishing and funding the Medical Genetics activities at GSK.

References

1. World Health Organization. Global surveillance, prevention and control of chronic respiratory diseases: a comprehensive approach, 2007. Available from: http://www.who.int/gard/publications/GARD_Manual/en/index.html
2. Lopez AD, Shibuya K, Rao C, Mathers CD, Hansell AL, Held LS, Schmid V, Buist S. Chronic obstructive pulmonary disease: current burden and future projections. *Eur Respir J* 2006;27:397–412.
3. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med* 2006;3:e442.
4. Weiss ST. Lung function and airway diseases. *Nat Genet* 2010;42:14–16.
5. McCarthy MI. Exploring the unknown: assumptions about allelic architecture and strategies for susceptibility variant discovery. *Genome Med* 2009;1:66.
6. Talmud PJ, Hingorani AD, Cooper JA, Marmot MG, Brunner EJ, Kumari M, Kivimaki M, Humphries SE. Utility of genetic and non-genetic risk factors in prediction of type 2 diabetes: Whitehall II prospective cohort study. *BMJ* 2010;340:b4838.
7. Cosio BG, Agusti A. Update in chronic obstructive pulmonary disease 2009. *Am J Respir Crit Care Med* 2010;181:655–660.
8. Hancock DB, Eijgelsheim M, Wilk JB, Gharib SA, Loehr LR, Marciante KD, Franceschini N, van Durme YM, Chen TH, Barr RG, *et al*. Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat Genet* 2010;42:45–52.
9. Repapi E, Sayers I, Wain LV, Burton PR, Johnson T, Obeidat M, Zhao JH, Ramasamy A, Zhai G, Vitart V, *et al*. Genome-wide association

study identifies five loci associated with lung function. *Nat Genet* 2010; 42:36–44.
10. Wilk JB, Chen TH, Gottlieb DJ, Walter RE, Nagle MW, Brandler BJ, Myers RH, Borecki IB, Silverman EK, Weiss ST, *et al.* A genome-wide association study of pulmonary function measures in the Framingham Heart Study. *PLoS Genet* 2009;5:e1000429.
11. Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, Feng S, Hersh CP, Bakke P, Gulsvik A, *et al.* A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet* 2009;5:e1000421.
12. Cho MH, Boutaoui N, Klanderman BJ, Sylvia JS, Ziniti JP, Hersh CP, DeMeo DL, Hunninghake GM, Litonjua AA, Sparrow D, *et al.* Variants in *FAM13A* are associated with chronic obstructive pulmonary disease. *Nat Genet* 2010;42:200–202.
13. Soler Artigas M, Wain L.V, Repapi E, Obeidat M, Sayers I, Hall I.P, Tobin M.D; SpiroMeta Consortium. Five new loci associated with lung function and their joint effect on lung function and COPD risk [abstract]. Presented at the European Mathematical Genetics Meeting, Oxford, 2010
14. Soler Artigas M, Wain L.V, Obeidat M, Repapi E, Sayers I, Hall I.P, Tobin M.D; SpiroMeta Consortium. New loci associated with lung function and chronic obstructive pulmonary disease [abstract]. Presented at the International Genetic Epidemiology Society Meeting, Boston, 2010
15. Hankinson JL, Crapo RO, Jensen RL. Spirometric reference values for the 6-s FVC maneuver. *Chest* 2003;124:1805–1811.
16. Hankinson JL, Kawut SM, Shahar E, Smith LJ, Stukovsky KH, Barr RG. Performance of American Thoracic Society–recommended spirometry reference values in a multiethnic sample of adults: the Multi-ethnic Study of Atherosclerosis (MESA) Lung Study. *Chest* 2010;137:138–145.
17. Global Initiative for Chronic Obstructive Lung Disease (GOLD). Global strategy for the diagnosis, management and prevention of COPD, 2006. Available from http://www.goldCOPD.org
18. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med* 1999;159:179–187.
19. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 2003; 33:177–182.
20. Van Durme YM, Eijgelsheim M, Joos GF, Hofman A, Uitterlinden AG, Brusselle GG, Stricker BH. Hedgehog-interacting protein is a COPD susceptibility gene: the Rotterdam Study. *Eur Respir J* 2010;36:89–95.
21. Liu JZ, Tozzi F, Waterworth DM, Pillai SG, Muglia P, Middleton L, Berrettini W, Knouff CW, Yuan X, Waeber G, *et al.* Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat Genet* 2010;42:436–440.
22. Kohansal R, Martinez-Camblor P, Agusti A, Buist AS, Mannino DM, Soriano JB. The natural history of chronic airflow obstruction revisited: an analysis of the Framingham offspring cohort. *Am J Respir Crit Care Med* 2009;180:3–10.
23. Hayes JD, Flanagan JU, Jowsey IR. Glutathione transferases. *Annu Rev Pharmacol Toxicol* 2005;45:51–88.
24. Weigt C, Gaertner A, Wegner A, Korte H, Meyer HE. Occurrence of an actin-inserting domain in tensin. *J Mol Biol* 1992;227:593–595.
25. Chen H, Duncan IC, Bozorgchami H, Lo SH. Tensin1 and a previously undocumented family member, tensin2, positively regulate cell migration. *Proc Natl Acad Sci USA* 2002;99:733–738.
26. Bayer H, Muller T, Myrtek D, Sorichter S, Ziegenhagen M, Norgauer J, Zissel G, Idzko M. Serotoninergic receptors on human airway epithelial cells. *Am J Respir Cell Mol Biol* 2007;36:85–93.
27. Miller L-AD, Wert SE, Clark JC, Xu Y, Perl A-KT, Whitsett JA. Role of Sonic hedgehog in patterning of tracheal–bronchial cartilage and the peripheral lung. *Dev Dyn* 2004;231:57–71.
28. Loos RJ, Lindgren CM, Li S, Wheeler E, Zhao JH, Prokopenko I, Inouye M, Freathy RM, Attwood AP, Beckmann JS, *et al.* Common variants near *MC4R* are associated with fat mass, weight and risk of obesity. *Nat Genet* 2008;40:768–775.
29. Fehrenbach H, Kasper M, Tschernig T, Shearman MS, Schuh D, Muller M. Receptor for advanced glycation endproducts (RAGE) exhibits highly differential cellular and subcellular localisation in rat and human lung. *Cell Mol Biol* 1998;44:1147–1157.
30. Gaens KH, Ferreira I, van der Kallen CJ, van Greevenbroek MM, Blaak EE, Feskens EJ, Dekker JM, Nijpels G, Heine RJ, 't Hart LM, *et al.* Association of polymorphism in the receptor for advanced glycation end products (RAGE) gene with circulating RAGE levels. *J Clin Endocrinol Metab* 2009;94:5174–5180.
31. Konishi K, Gibson KF, Lindell KO, Richards TJ, Zhang Y, Dhir R, Bisceglia M, Gilbert S, Yousem SA, Song JW, *et al.* Gene expression profiles of acute exacerbations of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2009;180:167–175.
32. Chen H, Herndon ME, Lawler J. The cell biology of thrombospondin-1. *Matrix biology: journal of the International Society for Matrix Biology* 2000;19:597–614.
33. Johannessen A, Omenaas ER, Bakke PS, Gulsvik A. Implications of reversibility testing on prevalence and risk factors for chronic obstructive pulmonary disease: a community study. *Thorax* 2005;60:842–847.
34. Anderson CA, Soranzo N, Zeggini E, Barrett JC. Synthetic associations are unlikely to account for many common disease genome-wide association signals. *PLoS Biol* 2011;9:e1000580.
35. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, Willer CJ, Jackson AU, Vedantam S, Raychaudhuri S, Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 2010;467:832–838.