

# A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13

Michael H. Cho<sup>1,2,\*</sup>, Peter J. Castaldi<sup>1</sup>, Emily S. Wan<sup>1,2</sup>, Mateusz Siedlinski<sup>1</sup>, Craig P. Hersh<sup>1,2</sup>, Dawn L. Demeo<sup>1,2</sup>, Blanca E. Himes<sup>1</sup>, Jody S. Sylvia<sup>1</sup>, Barbara J. Klanderman<sup>1</sup>, John P. Ziniti<sup>1</sup>, Christoph Lange<sup>3</sup>, Augusto A. Litonjua<sup>1,2</sup>, David Sparrow<sup>4,5,6</sup>, Elizabeth A. Regan<sup>7</sup>, Barry J. Make<sup>7</sup>, John E. Hokanson<sup>8</sup>, Tanda Murray<sup>9</sup>, Jacqueline B. Hetmanski<sup>9</sup>, Sreekumar G. Pillai<sup>10,†</sup>, Xiangyang Kong<sup>11</sup>, Wayne H. Anderson<sup>10</sup>, Ruth Tal-Singer<sup>11</sup>, David A. Lomas<sup>12</sup>, Harvey O. Coxson<sup>13</sup>, Lisa D. Edwards<sup>10</sup>, William MacNee<sup>14</sup>, Jørgen Vestbo<sup>15</sup>, Julie C. Yates<sup>10</sup>, Alvar Agusti<sup>16</sup>, Peter M.A. Calverley<sup>17</sup>, Bartolome Celli<sup>2</sup>, Courtney Crim<sup>10</sup>, Stephen Rennard<sup>18</sup>, Emiel Wouters<sup>19</sup>, Per Bakke<sup>20</sup>, Amund Gulsvik<sup>20</sup>, James D. Crapo<sup>7</sup>, Terri H. Beaty<sup>9</sup>, Edwin K. Silverman<sup>1,2</sup> and on behalf of the ICGN, ECLIPSE, and COPD Gene Investigators

<sup>1</sup>Channing Laboratory and <sup>2</sup>Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham & Women's Hospital, Boston, MA, USA, <sup>3</sup>Harvard School of Public Health, Boston, MA, USA, <sup>4</sup>School of Public Health and <sup>5</sup>School of Medicine, Boston University, Boston, MA, USA, <sup>6</sup>Veterans Administration Boston Healthcare System, Boston, MA, USA, <sup>7</sup>National Jewish Health, Denver, CO, USA, <sup>8</sup>Department of Epidemiology, Colorado School of Public Health, University of Colorado Denver, Aurora, CO, USA, <sup>9</sup>Johns Hopkins School of Public Health, Baltimore, MD, USA, <sup>10</sup>GlaxoSmithKline Research and Development, Research Triangle Park, NC, USA, <sup>11</sup>King of Prussia, PA, USA, <sup>12</sup>Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK, <sup>13</sup>University of British Columbia Department of Radiology and James Hogg Research Centre, Institute for Heart and Lung Health, Vancouver, BC, Canada, <sup>14</sup>University of Edinburgh, Edinburgh, UK, <sup>15</sup>Hvidovre Hospital, Hvidovre, Denmark, <sup>16</sup>Thorax Institute, Hospital Clinic IDIBAPS, Universitat Pompeu-Fabra (Barcelona) and CIBER Enfermedades Respiratorias (CIBERES), Fundación Caubet-Cimera, Mallorca, Spain, <sup>17</sup>University Hospital Aintree, Liverpool, UK, <sup>18</sup>University of Nebraska, Omaha, NE, USA, <sup>19</sup>University of Maastricht, Maastricht, the Netherlands and <sup>20</sup>Haukeland University Hospital and Institute of Medicine, University of Bergen, Bergen, Norway

Received July 20, 2011; Revised October 26, 2011; Accepted November 7, 2011

The genetic risk factors for chronic obstructive pulmonary disease (COPD) are still largely unknown. To date, genome-wide association studies (GWASs) of limited size have identified several novel risk loci for COPD at *CHRNA3/CHRNA5/IREB2*, *HHIP* and *FAM13A*; additional loci may be identified through larger studies. We performed a GWAS using a total of 3499 cases and 1922 control subjects from four cohorts: the Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE); the Normative Aging Study (NAS) and National Emphysema Treatment Trial (NETT); Bergen, Norway (GenKOLS); and the COPD Gene study. Genotyping was performed on Illumina platforms with additional markers imputed using 1000 Genomes data; results were summarized using fixed-effect meta-analysis. We identified a new genome-wide significant locus on chromosome 19q13 (rs7937, OR = 0.74,  $P = 2.9 \times 10^{-9}$ ). Genotyping this single

\*To whom correspondence should be addressed at: Channing Laboratory, Brigham & Women's Hospital, 181 Longwood Avenue, Boston, MA 02115, USA. Tel: +1 6175250897; Fax: +1 6175250958; Email: remhc@channing.harvard.edu or michael.cho@channing.harvard.edu

†Present address: Hoffman La Roche, Nutley, NJ, USA.

nucleotide polymorphism (SNP) and another nearby SNP in linkage disequilibrium (rs2604894) in 2859 subjects from the family-based International COPD Genetics Network study (ICGN) demonstrated supportive evidence for association for COPD ( $P = 0.28$  and  $0.11$  for rs7937 and rs2604894), pre-bronchodilator FEV<sub>1</sub> ( $P = 0.08$  and  $0.04$ ) and severe (GOLD 3&4) COPD ( $P = 0.09$  and  $0.017$ ). This region includes *RAB4B*, *EGLN2*, *MIA* and *CYP2A6*, and has previously been identified in association with cigarette smoking behavior.

## INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is defined as airflow limitation that is not fully reversible and is usually caused by exposure to noxious particles or gases—predominantly cigarette smoking, though other exposures such as biomass fuels are an important cause worldwide (1). COPD does not reverse with smoking cessation, and despite efforts to curtail cigarette smoking, COPD is a leading and increasing cause of morbidity and mortality. Worldwide, it is projected to rise to rank fifth in disease burden by 2020 (1), and in the USA, it now ranks as the third leading cause of death (2). The development of COPD among smokers is not uniform; a minority of smokers develops the disease (3), and lung function response to similar levels of cigarette smoke exposure varies greatly (4). Numerous studies have demonstrated a genetic component to COPD and smoking-related changes in lung function (5–8). However, the results of many small candidate gene studies have been inconsistent (9,10). Genome-wide association studies (GWASs) of COPD (11–13) have, to date, identified three susceptibility loci that have been well replicated (14–21). We report the results of a follow-up GWAS in four cohorts that identify a new COPD susceptibility locus.

## RESULTS

Baseline characteristics for subjects from the four cohorts: the Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE); Normative Aging Study (NAS) and National Emphysema Treatment Trial (NETT); Bergen, Norway COPD Cohort (GenKOLS); and the COPD-Gene study (first 1000 subjects) are shown in Table 1. Fixed-effects meta-analysis across the four cohorts for case–control status showed no evidence of substantial deviation from the null ( $\lambda = 1.01$ , Fig. 1). The most significantly associated single nucleotide polymorphisms (SNPs) were in the previously identified locus on chromosome 4 in *FAM13A* (Table 2). However, a new genome-wide significant locus was identified on chromosome 19q13. The top  $P$ -value was at rs7937 (OR = 0.74,  $P$ -value =  $2.88 \times 10^{-9}$ ). There was some evidence of heterogeneity ( $P$ -value = 0.15 for Cochrane's  $Q$ ,  $I^2 = 43.8$ ); however, similar results were found using a modified random-effects model (22) ( $P = 3.36 \times 10^{-9}$ ).

We genotyped rs7937 and another nearby genome-wide significant SNP in linkage disequilibrium (LD) (rs2604894,  $r^2 = 0.74$ ) in 983 probands and 1876 siblings from the family-based International COPD Genetics Network study (ICGN).  $P$ -values for the COPD affection status for the same risk

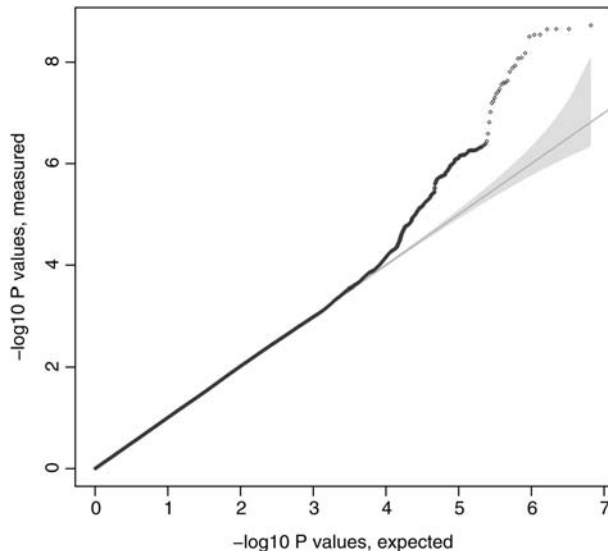
allele were 0.28 and 0.11 for rs7937 and rs2604894, respectively. More significant associations were demonstrated for pre-bronchodilator FEV<sub>1</sub> ( $P = 0.08$  and  $0.04$ , respectively) and after limiting cases to Global Initiative for Chronic Obstructive Lung Disease (GOLD) severity classifications 3 and 4 COPD ( $P = 0.09$  and  $0.017$ , respectively).

The 19q13 locus includes the genes *RAB4B*, *EGLN2* and *CYP2A6*, and SNPs in this locus have recently been identified in a large GWAS of smoking behavior (23,24) as associated with average number of cigarettes smoked per day; rs7937 was the second-ranked association to cigarettes per day at this locus in the study by Thorgeirsson *et al.* (23), with a combined  $P$ -value of  $2.4 \times 10^{-9}$ . To explore whether our findings were due to associations with cigarette smoking behavior, we examined the relationship of rs7937 and rs2604894 with both pack-years and average number of cigarettes smoked per day separately in the case and control groups. In none of these analyses were any significant associations found; the strongest association ( $P = 0.27$ ) was with pack-years of smoking in controls. These results are consistent with the results reported for rs7937 in the Bergen, Norway GenKOLS cohort, as this cohort was included in the GWAS by Thorgeirsson *et al.* (23). While these main effects were not significant, we also examined whether there was any evidence of a gene-by-cigarettes per day or pack-years interaction; none of these analyses were significant ( $P > 0.1$ ).

At chromosome 19q13, and at the three previously demonstrated genome-wide associated loci (11–13) (4q22—*FAM13A*; 4q31—*HHIP*; and 15q25—*CHRNA3/CHRNA5/IREB2*), we used the imputed marker data from the 1000 Genomes study to attempt to further refine previously determined association signals (Table 2 and Fig 2). At 4q22 and 4q31, imputed SNPs in high LD ( $r^2$  of 1.0 and 0.9, respectively) with the most significant genotyped association had only marginally smaller  $P$ -values. At 15q25, the top hit was rs11858836; this SNP is in strong LD with the previously reported rs8034191 and rs1051730 ( $r^2$  of  $\sim 0.75$ – $0.8$ ), both part of locus 1 of Saccone *et al.* (25), but also in moderate LD ( $r^2 \sim 0.4$ ) with rs13180, our previously reported most significant SNP in this region. We also attempted to assess whether any known non-synonymous SNPs could account for these association signals. At only one locus, 15q25, was there any non-synonymous SNP at an  $r^2$  of  $> 0.5$  with the most significant SNP in the region; this was the *CHRNA5* SNP rs16969968, which has been described as part of locus 1 above. To evaluate the potential effect of rarer SNPs—those with low  $r^2$  but high  $D'$ —we additionally evaluated SNPs with  $D' > 0.8$  within 500 kb of the top four loci. Two other imputed SNPs had  $D' \sim 1$  with rs7937 at 19q13 and nominal associations with COPD: rs1801272 in *CYP2A6*

**Table 1.** Baseline characteristics (mean  $\pm$  SD or percentage)

	COPDGene		ECLIPSE		NETT/NAS		GenKOLS	
	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls
<i>n</i>	499	501	1764	178	373	435	863	808
Age	64.77 (8.12)	60.2 (8.66)	63.63 (7.1)	57.48 (9.44)	67.47 (5.78)	69.8 (7.49)	65.53 (10.03)	55.62 (9.71)
Pack-years	54.76 (26.69)	38.87 (21.07)	50.29 (27.42)	32.11 (24.84)	66.43 (30.68)	40.66 (27.85)	31.98 (18.46)	19.66 (13.58)
Average cigarettes per day	27.58 (11.76)	24.89 (11.18)	25.54 (12.39)	21.87 (11.35)	32.55 (13.47)	29.27 (14.33)	15.6 (7.71)	13.82 (7.44)
FEV <sub>1</sub> , % predicted	48.73 (18.41)	97.98 (11.32)	47.63 (15.62)	107.83 (13.56)	28.12 (7.38)	99.97 (13.2)	50.63 (17.44)	94.91 (9.24)
Sex (% male)	49.5	50.1	67.0	57.9	63.8	100	60.1	50.1

**Figure 1.** The Quantile–quantile plot for the four-cohort meta-analysis including 1000 Genomes project imputed data, after adjustment for genetic ancestry using principal components.

[ $P = 0.03$ , minor allele frequency (MAF) = 0.04, imputation  $r^2$  0.3–0.4 using HapMap2 data] previously described in this cohort as associated with lifetime average cigarettes per day (26), and rs36012476 in *ADCK4* ( $P = 0.02$ , MAF = 0.07, and imputation  $r^2$  0.6–0.7). Inclusion of these SNPs only modestly attenuated the signal for rs7937 ( $P$  after adjustment =  $8.3 \times 10^{-8}$ ). In addition to single SNP analyses, we also performed analyses conditioning on the top SNP in each of the four GWAS regions, using markers 250 kb on either side and a threshold of  $P < \sim 5 \times 10^{-4}$  to reflect an approximate adjustment for a 500 kb interval flanking the most significant SNP. None of the loci demonstrated evidence of other independent SNP effects.

Finally, we attempted to estimate the relative contribution of these four risk loci to the sibling relative risk of COPD. We used odds ratios aggregated over the four cohorts and assumed independence of the four loci under a standard multiplicative model; acknowledging that the resulting estimates using the discovery cohorts are likely to be inflated. These four loci accounted for  $\sim 5\%$  of the total variance explained using either pseudo- $R^2$  or a liability threshold model. Alternatively, using estimates of sibling relative risk from family-based studies of 2–3 (5,6), these four loci account for  $\sim 8$ –15% of the sibling relative risk.

## DISCUSSION

While a familial component to COPD has been recognized for at least the last 40 years (27)—aside from the nearly-contemporaneous identification of alpha-1 antitrypsin deficiency (28)—the genetic determinants of COPD have been generally elusive. Our analysis adds an additional cohort to three prior GWASs and identifies a new genome-wide significant locus associated with COPD susceptibility at chromosome 19q13.

The 19q13 locus has demonstrated clear associations with smoking behavior, and has been shown to be associated with cigarettes per day in two large meta-analytic GWASs (23,24). An association operating through cigarette smoking is clearly the most parsimonious and most likely explanation. However, we cannot rule out the possibility that other mechanisms may also be responsible for the association to COPD that we observed. At 15q25, mediation approaches suggest that pack-years explain only  $\sim 1/4$  of the association between these SNPs and COPD (29). Another gene at this locus, *IREB2*, was identified independently in part through gene expression and genetic association (30) as a candidate gene for COPD. In our data set, we were unable to show a significant association between rs7937 and cigarettes per day or pack-years in any of our study populations, and our finding persisted after inclusion of a non-synonymous SNP associated with cigarettes per day in this cohort. The 19q13 locus includes several other genes of potential interest expressed in developing animal or human lung: *RAB4B* (31), *MIA* (32) and *EGLN2* (33). *LTBP4* is  $\sim 150$  kb away, and variants in this gene have been associated with functional outcomes in subjects with emphysema (34), and disruption of this gene leads to emphysema in mice (35) and in humans, as part of the Urban-Rifkin-Davis Syndrome (OMIM #613177) (36).

While future studies (both statistical analyses, such as causal modeling and mediation analysis, and functional studies) may help determine whether other genes at this locus play a role in the 19q13 association, it is highly likely that this association occurs through cigarette smoking. *CYP2A6* variants have additionally been associated with another major smoking-related disease, lung cancer (37). *CYP2A6* and, to a lesser extent, *CYP2B6* are involved in nicotine metabolism (38,39); the most significant associations in these studies are in LD with sequence variants that have been shown to reduce *CYP2A6* enzyme activity, and rs7937 was associated with the levels of the major nicotine metabolite cotinine in a subset of subjects from the European Network for Genetic and Genomic Epidemiology (23). While we were able to identify a nominally significant association of cigarettes per day with a non-



Table 2. Meta-analysis results

Locus	SNP	Minor allele	MAF	COPDGene OR (95% CI)	P-value	ECLIPSE OR (95% CI)	P-value	NETT/NAS OR (95% CI)	P-value	GenKOLS OR (95% CI)	P-value	Overall OR (95% CI)	P-value	I <sup>2</sup>	P-value (Q)
4q22	rs1964516	C	0.49	0.78 <sup>a</sup> (0.64–0.95)	1.39 × 10 <sup>−2</sup>	0.72 <sup>a</sup> (0.56–0.91)	7.37 × 10 <sup>−3</sup>	0.62 <sup>a</sup> (0.49–0.8)	2.27 × 10 <sup>−4</sup>	0.76 <sup>a</sup> (0.64–0.89)	7.58 × 10 <sup>−4</sup>	0.73 (0.66–0.81)	1.88 × 10 <sup>−9</sup>	0	0.56
	rs7671167	C	0.49	0.78 (0.64–0.95)	1.38 × 10 <sup>−2</sup>	0.72 (0.57–0.92)	7.75 × 10 <sup>−3</sup>	0.63 (0.49–0.81)	2.69 × 10 <sup>−4</sup>	0.76 (0.64–0.89)	7.67 × 10 <sup>−4</sup>	0.73 (0.66–0.81)	2.22 × 10 <sup>−9</sup>	0	0.58
19q13	rs7937	C	0.44	0.64 (0.52–0.78)	1.06 × 10 <sup>−5</sup>	0.72 (0.57–0.92)	9.48 × 10 <sup>−3</sup>	0.68 (0.54–0.86)	1.35 × 10 <sup>−3</sup>	0.85 (0.72–0.99)	4.06 × 10 <sup>−2</sup>	0.73 <sup>b</sup> (0.63–0.83)	2.88 × 10 <sup>−9</sup>	44	0.15
	rs2604894	A	0.44	0.66 <sup>b</sup> (0.53–0.81)	9.28 × 10 <sup>−5</sup>	0.72 <sup>a</sup> (0.56–0.92)	9.38 × 10 <sup>−3</sup>	0.68 <sup>a</sup> (0.53–0.88)	2.67 × 10 <sup>−3</sup>	0.85 <sup>a</sup> (0.72–1)	5.25 × 10 <sup>−2</sup>	0.74 <sup>b</sup> (0.65–0.84)	3.41 × 10 <sup>−8</sup>	30	0.23
4q31	rs13141641	C	0.43	0.91 <sup>a</sup> (0.75–1.1)	3.35 × 10 <sup>−1</sup>	0.7 <sup>a</sup> (0.55–0.89)	3.00 × 10 <sup>−3</sup>	0.62 <sup>a</sup> (0.48–0.79)	1.39 × 10 <sup>−4</sup>	0.78 <sup>a</sup> (0.66–0.92)	3.06 × 10 <sup>−3</sup>	0.76 <sup>b</sup> (0.65–0.88)	2.56 × 10 <sup>−7</sup>	54	0.09
	rs13118928	G	0.40	0.92 <sup>a</sup> (0.76–1.12)	3.95 × 10 <sup>−1</sup>	0.74 (0.58–0.93)	1.11 × 10 <sup>−2</sup>	0.6 (0.47–0.78)	8.04 × 10 <sup>−5</sup>	0.78 (0.66–0.92)	2.45 × 10 <sup>−3</sup>	0.76 <sup>b</sup> (0.65–0.89)	5.97 × 10 <sup>−7</sup>	57	0.07
15q25	rs11858836	A	0.37	1.25 <sup>a</sup> (1.02–1.53)	2.98 × 10 <sup>−2</sup>	1.05 <sup>a</sup> (0.82–1.34)	7.13 × 10 <sup>−1</sup>	1.49 <sup>a</sup> (1.16–1.92)	1.98 × 10 <sup>−3</sup>	1.38 <sup>a</sup> (1.16–1.64)	2.33 × 10 <sup>−4</sup>	1.29 <sup>b</sup> (1.13–1.47)	1.35 × 10 <sup>−6</sup>	35	0.20
	rs13180	C	0.36	0.78 (0.64–0.95)	1.52 × 10 <sup>−2</sup>	0.78 (0.61–0.99)	4.26 × 10 <sup>−2</sup>	0.74 (0.58–0.95)	1.75 × 10 <sup>−2</sup>	0.79 (0.67–0.94)	7.59 × 10 <sup>−3</sup>	0.78 (0.7–0.86)	2.02 × 10 <sup>−6</sup>	0	0.98

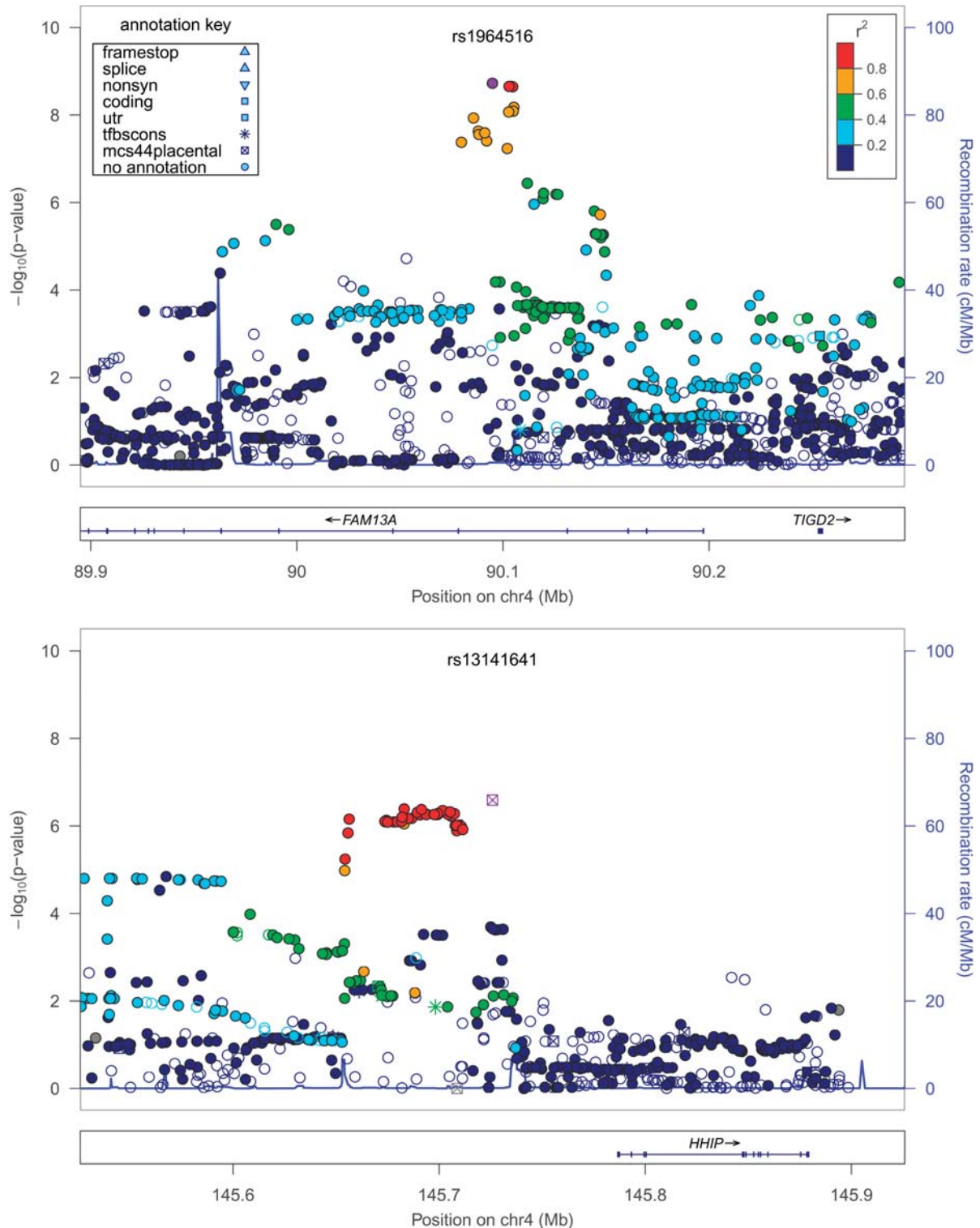
The top result at each locus as well as additional genotyped SNPs at that locus is shown.  
<sup>a</sup>Imputed genotypes; note that imputed SNPs in this table had imputation quality  $r^2 > 0.9$ .  
<sup>b</sup>Effect estimates given for random-effects model.

synonymous SNP in *CYP2A6* in our COPD cohorts (26), *CYP2A6* activity is affected by a number of other demonstrated functional variants—both common and rare (40–43). While we failed to demonstrate an association with cigarette smoking behavior and rs7937 within our case or control groups, both the 15q25 and 19q13 loci have been associated with cotinine levels in other studies that did not find an association with cigarettes per day (44,45), suggesting that standard measures of smoking behavior are incomplete. Issues with measuring the complexity of cigarette smoke exposure such as recall bias (46), as well as differential intensity of smoking per cigarette and effects of cigarette metabolism, may be responsible for the lack of association with pack-years or cigarettes per day in our case and control groups.

The identification of this 19q13 COPD locus, in conjunction with 15q25, highlights the critical contribution of variants affecting the major behavioral risk factor for COPD, cigarette smoking. This is in contrast to coronary artery disease, where despite the fact that cigarette smoking is a major risk factor (47), none of the identified genome-wide association loci to date has identified variants known to affect smoking, and a minority of the identified loci have been associated with traditional risk factors (48). Ongoing cigarette smoking causes accelerated lung function decline; conversely, smoking cessation attenuates this decline, improves respiratory symptoms and reduces overall mortality (49–52). Decreasing cigarette consumption is essential to reducing the risk of COPD, and though a subset of our subjects carry lower risk alleles, our data do not suggest that there are smokers unlikely to benefit from smoking cessation. Whether genetic testing may aid a subset of more susceptible individuals in smoking cessation efforts is unclear (53,54).

While our GWAS is the largest reported to date for COPD, our sample size is still substantially smaller—by a factor of 10 or more—than some other complex diseases; a recent coronary artery disease GWAS included 22 233 cases (48). Thus, if the genetic architecture of COPD is similar to other complex common diseases, our study is underpowered to detect many of the likely other common variants contributing to COPD susceptibility. This lack of power may also explain why we were unable to discern independent associations at the 15q25 locus, as others have demonstrated (25); our family-based replication results may have also suffered from power limitations, or, less likely, reflect subtle stratification in our original case–control analysis (55,56). Our study also did not directly assess for rare genetic variation, which may be an important contributor to COPD susceptibility (57), nor did we assess for copy number variation. Finally, we did not attempt to address issues of COPD heterogeneity, for example, through radiographic phenotypes (58).

Our GWAS increases the number of genome-wide significant loci associated with COPD to four. Our analysis at these risk loci, using publicly available dense sequencing data and imputation, helps refine these signals for further study and replication. These loci account for only a small fraction of the observed effect of all genetic variants in COPD risk, and we anticipate analysis of other types of genetic variation (rare variants, copy number variants) and perhaps more importantly—future collaborations to increase the available sample size (59)—will expand this list of genetic loci and improve our understanding of COPD susceptibility.



**Figure 2.** Association plots at genome-wide significant loci.  $R^2$  values are given in relation to the most significant, labeled and highlighted SNP. Plots created using LocusZoom (83).

## MATERIALS AND METHODS

Genotyping methods and study descriptions for three data sets: the ECLIPSE; NAS and NETT; and GenKOLS have been

described previously (11,13,60–64). All cohorts were of self-described European white ancestry and genotyped on Illumina platforms (Human Hap550 or Quad610). Quality-control procedures included tests for subject missingness, discordances,

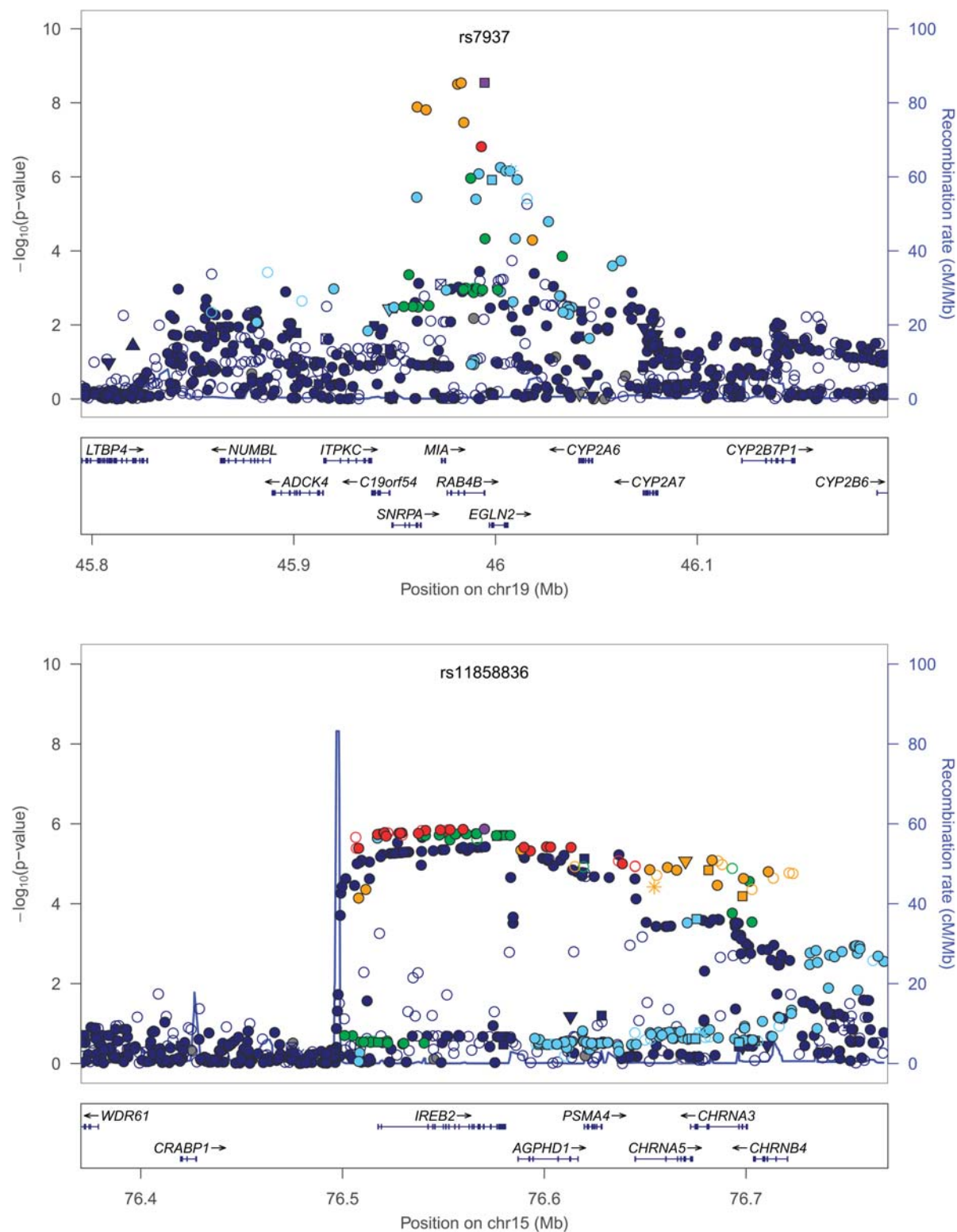


Figure 2. Continued

relatedness and sex; and marker missingness, discordances, singletons and Hardy-Weinberg equilibrium.

The study protocol for COPDGene (NCT00608764) has been described previously (65). Briefly, COPDGene is a multi-center

genetic and epidemiologic investigation to study COPD and other smoking-related lung diseases. Participants completed a detailed protocol, including questionnaires, pre- and post-bronchodilator spirometry, high-resolution CT scanning of the



chest and blood samples for genotyping. Samples from self-described European whites were genotyped at the Center for Inherited Disease Research (CIDR) at Johns Hopkins University using the Illumina Omni1 Quad platform. COPDGene genotyping underwent a similar quality-control procedure: subjects were screened for missing call rates > 1%, relatedness by estimated identity-by-descent > 0.125, sex discrepancies, and inbreeding coefficients > 0.2. Markers were screened for missingness > 2%, minor allele frequency < 1%, and deviation from Hardy–Weinberg at  $P < 1 \times 10^{-5}$ . Out of an initial set of 1006 COPDGene subjects, only one was unable to be genotyped successfully; none of the others failed on quality-control criteria. Five additional subjects were excluded based on the presence of lung parenchymal abnormalities other than emphysema after chest CT scan review. The average per-subject genotype missing rate was 0.05%, with a maximum missing rate of 0.96%. Details on the COPDGene subjects, as well as the subjects from the other three cohorts passing initial quality-control before removal of principal component outliers are shown in Table 1.

Of the 986 763 autosomal markers in the COPDGene study successfully genotyped, 797 983 markers remained after quality-control, with >95% of these exclusions on the basis of low minor allele frequency (<1%). Combining these markers with those from the other three cohorts (ECLIPSE, NETT/NAS and GenKOLS), a total of 296 201 markers were shared among all cohorts. After imputation from the 1000 Genomes study, this number increased to ~6.1 million SNPs. Analysis for and control of population stratification were also performed via principal components using EIGEN-SOFT2.0 (66) as previously described (13) for the ECLIPSE, NETT, NAS and GenKOLS cohorts. For COPDGene, pruned markers with a LD cutoff of 0.12 were chosen from those with a MAF > 0.05, Illumina GC scores of >0.8 and present in HapMap CEU subjects. Outliers were removed over five iterations with deviations beyond 6 standard deviations, along the top 10 principal components. Significant principal components were defined using the Tracy–Widom statistic (67) as previously described (13). After removal of additional subjects with cryptic relatedness between cohorts and outliers by genetic ancestry based on principal components, 3456 cases and 1908 controls (99% of subjects) remained in the analysis. Ten ECLIPSE subjects were found to have borderline FEV<sub>1</sub>/FVC ratios but were retained in the analysis.

Genotype imputation within each study was performed using MaCH 1.0.16 (68) using 100 rounds of iterations to estimate model parameters and CEU samples from HapMap2 (69) and the 1000 Genomes Project (70) (phased CEU data, March 2010) as reference populations. Markers with an imputation  $r^2 \leq 0.3$  were dropped from further analysis. Association analysis of SNPs with case–control status was performed in each cohort using logistic regression, adjusting for age, pack-years of cigarette smoking and genetic ancestry as summarized in the principal components. Imputed genotypes were analyzed in a similar manner, using SNP dosage data in PLINK 1.07 (71). Results were analyzed among the four cohorts using fixed-effect meta-analyses (72) using METAL (73) and in R 2.12 (www.r-project.org) with the meta-package. Genomic inflation factors (74) were calculated using GenABEL (75).  $I^2$  and Cochran's  $Q$  were

calculated to assess for heterogeneity; a secondary analysis using a modified random-effects model was also performed (22).

Replication genotyping was performed using the SEQUENOM MassARRAY MALDI-TOF mass spectrometer (Sequenom, San Diego, CA, USA). Details of the ICGN cohort used for replication have been previously described (13,58). Association analysis in ICGN was performed using PBAT 3.61 (76) using one-sided  $P$ -values for the same risk allele, adjusting for age and pack-years of smoking in the COPD affection status analysis, and age, pack-years, sex and height in the analysis of FEV<sub>1</sub>.

Analyses conditional on other SNPs were performed by extracting genotype dosage data (for imputed SNPs) or actual genotypes; logistic regression adjusting for significantly associated SNPs was performed in R 2.12. Calculation of the contribution of loci to sibling relative risk was obtained using overall odds ratios and minor allele frequencies (77,78). Estimation of the fraction of variance explained was calculated using logistic regression and Nagelkerke's pseudo- $R^2$  (79) as implemented in R Design package, and alternatively using the estimates based on a liability threshold model by So *et al.* (80) using a prevalence of COPD in smokers of ~10% (81).

LD was calculated for HapMap2 and 1000 Genomes CEU data using SNAP (82). Regional association plots were created using LocusZoom (83). All positions are given in reference to the Human March 2006 (NCBI36/hg18) assembly.

## AUTHOR CONTRIBUTIONS

Study design: E.K.S., M.H.C., A.A.L., D.S., S.G.P., X.K., W.H.A., R.T.S., D.A.L., P.B., A.G., T.H.B., E.A.R., B.J.M., J.E.H., J.D.C., H.O.C., L.D.E., W.M., J.V., J.C.Y., A.A., P.M.A.C., B.C., C.C., S.R., E.W.

Sample collection and phenotyping: A.A.L., D.S., S.G.P., X.K., W.H.A., R.T.S., D.A.L., P.B., A.G., E.A.R., B.J.M., J.D.C., E.K.S., C.P.H., D.L.D., H.O.C., L.D.E., W.M., J.V., J.C.Y., A.A., P.M.A.C., B.C., C.C., S.R., E.W.

Genotyping: B.J.K., M.H.C., S.G.P., X.K., T.H.B., T.M., J.B.H.

Informatics: M.H.C., J.S.S., J.P.Z., B.J.K., E.S.W., P.J.C., B.E.H.

Statistical analysis: M.H.C., C.L., T.H.B., T.M., J.B.H., M.S., E.K.S.

Manuscript writing: M.H.C., E.K.S.

## ACKNOWLEDGEMENTS

The members of the COPDGene® study group include: Ann Arbor VA: Jeffrey Curtis, MD (PI), Ella Kazerooni, MD (RAD). Baylor College of Medicine, Houston, TX: Nicola Hanania, MD, MS (PI), Philip Alapat, MD, Venkata Bandi, MD, Kalpalatha Guntupalli, MD, Elizabeth Guy, MD, Antara Mallampalli, MD, Charles Trinh, MD (RAD), Mustafa Atik, MD. Brigham and Women's Hospital, Boston, MA: Dawn DeMeo, MD, MPH (Co-PI), Craig Hersh, MD, MPH (Co-PI), George Washko, MD, Francine Jacobson, MD, MPH (RAD). Columbia University,

New York, NY: R. Graham Barr, MD, DrPH (PI), Byron Thomashow, MD, John Austin, MD (RAD).

Duke University Medical Center, Durham, NC: Neil MacIntyre, Jr, MD (PI), Lacey Washington, MD (RAD), H. Page McAdams, MD (RAD). Fallon Clinic, Worcester, MA: Richard Rosiello, MD (PI), Timothy Bresnahan, MD (RAD). Health Partners Research Foundation, Minneapolis, MN: Charlene McEvoy, MD, MPH (PI), Joseph Tashjian, MD (RAD). Johns Hopkins University, Baltimore, MD: Robert Wise, MD (PI), Nadia Hansel, MD, MPH, Robert Brown, MD (RAD), Gregory Diette, MD. Los Angeles Biomedical Research Institute at Harbor UCLA Medical Center, Los Angeles, CA: Richard Casaburi, MD (PI), Janos Porszasz, MD, PhD, Hans Fischer, MD, PhD (RAD), Matt Budoff, MD.

Michael E. DeBakey VAMC, Houston, TX: Amir Sharafkhaneh, MD (PI), Charles Trinh, MD (RAD), Hirani Kamal, MD, Roham Darvishi, MD. Minneapolis VA: Dennis Niewoehner, MD (PI), Tadashi Allen, MD (RAD), Quentin Anderson, MD (RAD), Kathryn Rice, MD. Morehouse School of Medicine, Atlanta, GA: Marilyn Foreman, MD, MS (PI), Gloria Westney, MD, MS, Eugene Berkowitz, MD, PhD (RAD). National Jewish Health, Denver, CO: Russell Bowler, MD, PhD (PI), Adam Friedlander, MD, David Lynch, MB (RAD), Joyce Schroeder, MD (RAD), John Newell, Jr., MD (RAD).

Temple University, Philadelphia, PA: Gerard Criner, MD (PI), Victor Kim, MD, Nathaniel Marchetti, DO, Aditi Satti, MD, A. James Mamary, MD, Robert Steiner, MD (RAD), Chandra Dass, MD (RAD). University of Alabama, Birmingham, AL: William Bailey, MD (PI), Mark Dransfield, MD (Co-PI), Hrudaya Nath, MD (RAD). University of California, San Diego, CA: Joe Ramsdell, MD (PI), Paul Friedman, MD (RAD).

University of Iowa, Iowa City, IA: Geoffrey McLennan, MD, PhD (PI), Edwin JR van Beek, MD, PhD (RAD), Brad Thompson, MD (RAD), Dwight Look, MD. University of Michigan, Ann Arbor, MI: Fernando Martinez, MD (PI), MeiLan Han, MD, Ella Kazerooni, MD (RAD). University of Minnesota, Minneapolis, MN: Christine Wendt, MD (PI), Tadashi Allen, MD (RAD). University of Pittsburgh, Pittsburgh, PA: Frank Sciurba, MD (PI), Joel Weissfeld, MD, MPH, Carl Fuhrman, MD (RAD), Jessica Bon, MD. University of Texas Health Science Center at San Antonio, San Antonio, TX: Antonio Anzueto, MD (PI), Sandra Adams, MD, Carlos Orozco, MD, Mario Ruiz, MD (RAD).

**Administrative Core:** James Crapo, MD (PI), Edwin Silverman, MD, PhD (PI), Barry Make, MD, Elizabeth Regan, MD, Sarah Moyle, MS, Douglas Stinson.

**Genetic Analysis Core:** Terri Beaty, PhD, Barbara Klanderma, PhD, Nan Laird, PhD, Christoph Lange, PhD, Michael Cho, MD, Stephanie Santorico, PhD, John Hokanson, MPH, PhD, Dawn DeMeo, MD, MPH, Nadia Hansel, MD, MPH, Craig Hersh, MD, MPH, Jacqueline Hetmanski, MS, Tanda Murray.

**Imaging Core:** David Lynch, MB, Joyce Schroeder, MD, John Newell, Jr, MD, John Reilly, MD, Harvey Coxson, PhD, Philip Judy, PhD, Eric Hoffman, PhD, George Washko, MD, Raul San Jose Estepar, PhD, James Ross, MSc, Rebecca Leek, Jordan Zach, Alex Kluiber, Jered Sieren, Heather Baumhauer, Verity McArthur, Dzimitry Kazlouski, Andrew Allen, Tanya Mann, Anastasia Rodionova.

**PFT QA Core, LDS Hospital, Salt Lake City, UT:** Robert Jensen, PhD.

**Biological Repository, Johns Hopkins University, Baltimore, MD:** Homayoon Farzadegan, PhD, Stacey Meyerer, Shivam Chandan, Samantha Bragan.

**Data Coordinating Center and Biostatistics, National Jewish Health, Denver, CO:** James Murphy, PhD (deceased), Douglas Everett, PhD, Carla Wilson, MS, Ruthie Knowles, Amber Powell, Joe Piccoli, Maura Robinson, Margaret Forbes, Martina Wamboldt.

**Epidemiology Core, University of Colorado School of Public Health, Denver, CO:** John Hokanson, MPH, PhD, Marci Sontag, PhD, Jennifer Black-Shinn, MPH, Gregory Kinney, MPH.

Principal investigators and centers participating in **ECLIPSE** include: Bulgaria: Y. Ivanov, Plevan; K. Kostov, Sofia. Canada: J. Bourbeau, Montreal; M. Fitzgerald, Vancouver; P. Hernández, Halifax; K. Killian, Hamilton; R. Levy, Vancouver; F. Maltais, Montreal; D. O'Donnell, Kingston. Czech Republic: J. Krepelka, Praha.

Denmark: J. Vestbo, Hvidovre. The Netherlands: E. Wouters, Horn. New Zealand: D. Quinn, Wellington. Norway: P. Bakke, Bergen, Slovenia: M. Kosnik, Golnik. Spain: A. Agusti, Jaume Saulea, Palma de Mallorca. Ukraine: Y. Feschenko, Kiev; V. Gavrisyuk, Kiev; L. Yashina, Kiev. UK: L. Yashina, W. MacNee, Edinburgh; D. Singh, Manchester; J. Wedzicha, London. USA: A. Anzueto, San Antonio, TX; S. Braman, Providence, RI; R. Casaburi, Torrance, CA; B. Celli, Boston, MA; G. Giessel, Richmond, VA; M. Gotfried, Phoenix, AZ; G. Greenwald, Rancho Mirage, CA; N. Hanania, Houston, TX; D. Mahler, Lebanon, NH; B. Make, Denver, CO; S. Rennard, Omaha, NE; C. Rochester, New Haven, CT; P. Scanlon, Rochester, MN; D. Schuller, Omaha, NE; F. Sciurba, Pittsburgh, PA; A. Sharafkhaneh, Houston, TX; T. Siler, St Charles, MO; E. Silverman, Boston, MA; A. Wanner, Miami, FL; R. Wise, Baltimore, MD; R. ZuWallack, Hartford, CT.

**Steering Committee:** H. Coxson (Canada), C. Crim (GlaxoSmithKline, USA), L. Edwards (GlaxoSmithKline, USA), D. Lomas (UK), W. MacNee (UK), E. Silverman (USA), R. Tal Singer (Co-chair, GlaxoSmithKline, USA), J. Vestbo (Co-chair, Denmark), J. Yates (GlaxoSmithKline, USA).

**Scientific Committee:** A. Agusti (Spain), P. Calverley (UK), B. Celli (USA), C. Crim (GlaxoSmithKline, USA), B. Miller (GlaxoSmithKline, USA), W. MacNee (Chair, UK), S. Rennard (USA), R. Tal-Singer (GlaxoSmithKline, USA), E. Wouters (The Netherlands), J. Yates (GlaxoSmithKline, USA).

The **National Emphysema Treatment Trial** was supported by the National Heart, Lung, and Blood Institute contracts N01HR76101, N01HR76102, N01HR76103, N01HR76104, N01HR76105, N01HR76106, N01HR76107, N01HR76108, N01HR76109, N01HR76110, N01HR76111, N01HR76112, N01HR76113, N01HR76114, N01HR76115, N01HR76116, N01HR76118 and N01HR76119.

Co-investigators in the NETT Genetics Ancillary Study also include J. Benditt, G. Criner, M. DeCamp, P. Diaz, M. Ginsburg, L. Kaiser, M. Katz, M. Krasna, N. MacIntyre, R. McKenna, F. Martinez, Z. Mosenifar, J. Reilly, A. Ries, P. Scanlon, F. Sciurba and J. Utz.



**International COPD Genetics Network (ICGN)** investigators: Edwin K. Silverman, Brigham & Women's Hospital, Boston, MA, USA; David A. Lomas, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK; Barry J. Make, National Jewish Medical and Research Center, Denver, CO, USA; Alvar Agusti and Jaume Sauleda, Hospital Universitari Son Dureta, Fundació Caubet-Cimera and Ciber Enfermedades Respiratorias, Spain; Peter M.A. Calverley, University of Liverpool, UK; Claudio F. Donner, Division of Pulmonary Disease, S. Maugeri Foundation, Veruno (NO), Italy; Robert D. Levy, University of British Columbia, Vancouver, Canada; Peter D. Paré, University of British Columbia, Vancouver, Canada; Stephen Rennard, Section of Pulmonary & Critical Care, University of Nebraska Medical Center, Omaha, NE, USA; Jørgen Vestbo, Department of Cardiology and Respiratory Medicine, Hvidovre Hospital, Copenhagen, Denmark; Emiel F.M. Wouters, University Hospital Maastricht, The Netherlands.

**Conflict of Interest statement.** R.T.-S., X.K., L.D.E., J.Y. and C.C. are current employees of GlaxoSmithKline. E.K.S. has received grant support from GlaxoSmithKline for studies of COPD genetics and honoraria and consulting fees from AstraZeneca. D.A.L. has received educational grants, fees for speaking and acts as a consultant for GlaxoSmithKline. H.O.C. has received grant support, honorarium, travel expenses, and has contract service agreements with GlaxoSmithKline or GlaxoSmithKline Canada; has a service agreement with Spiration Inc., has received grant support from a Canadian Institutes of Health—Industry (Wyeth) partnership grant, and has received a fee for speaking at a conference and related travel expenses from AstraZeneca (Australia). S.J.R. consulted or participated in advisory boards for: ABIM, Able Associates, Adelphi Research, Almirall, APT, Aradigm, Argenta, AstraZeneca, BI (ACCP), Biostrategies, BoomCom, Britnall and Nicolini, Capital Research, Chiesi, Clinical Advisors, CommonHealth, Complete Medical Group, Consult Complete, COPDForum, DataMonitor, Decision Resources, Defined Health, Dey, Dunn Group, Easton Associates, Enterprise Analysis, Equinox, Forest, Fulcrum, Gerson Lehman, GSK, Guidepoint, Hoffman LaRoche, IMS, Informed, Inspire, Insyght, KOL Connection, Leerink Swan, M. Pankove, MDRx Financial, MedaCorp, Medimmune, Mpex, Novartis, Nycomed, Oriel, Otsuka, Pearl, Pennside Partners, Pfizer, Pharma Ventures, Pharmaxis, Pick Research, Prescott, Price Waterhouse, Propagate, Pulmatrix, Pulmonary Reviews, Quadrant, Reckner Associates, Recruiting Resource, Reviews and Trends in COPD/Convergent Health Solutions, Roche, Saco, Schering, Schlesinger Medical, Scimed, Smith Research, Sudler and Hennessey, Talecris, Theravance, UBC, Uptake Medical, Vantage Point; has received lecture fees from AAAAI, Am Col Osteopathic Physicians, Asan Medical Center, ATS, AstraZeneca, California Soc Allergy, Convergent Health Solutions for Reviews and Trends in COPD, COPDFoundation, Creative Educational Concepts, Dey, Duke, France Foundation, Information TV, University of California-Los Angeles, Network for Continuing Education, Novartis, Nycomed, Otsuka, Pfizer, Sarasota Mem Hospital, Spanish Thoracic Society, University of Washington, University of Alabama-Birmingham, University of

Pittsburgh, University of British Columbia, University of California-Davis, VA Sioux Falls; and has received industry-sponsored grants from AstraZeneca, Biomarck, Centocor, GlaxoSmithKline, Mpex, Nabi, Novartis, Otsuka, Pfizer.

## FUNDING

This work was supported by US National Institutes of Health (NIH) grants R01 HL075478, R01 HL084323, P01 HL083069, P01 HL105339 and U01 HL089856 (E.K.S.); K12HL089990 and K08 HL097029 (M.H.C.); and U01 HL089897 (J.D.C.). M.S. is a recipient of a postdoctoral fellowship from the Niels Stensen Foundation. The National Emphysema Treatment Trial was supported by the National Heart, Lung, and Blood Institute, the Centers for Medicare and Medicaid Services and the Agency for Healthcare Research and Quality. The Normative Aging Study is supported by the Cooperative Studies Program/ERIC of the US Department of Veterans Affairs and is a component of the Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC). The Norway GenKOLS study (Genetics of Chronic Obstructive Lung Disease, GSK code RES11080), the ECLIPSE study (clinicaltrials.gov identifier NCT00292552; GSK code SCO104960) and the ICGN study are funded by GlaxoSmithKline. The COPDGene® project is also supported by the COPD Foundation through contributions made to an Industry Advisory Board comprised of AstraZeneca, Boehringer Ingelheim, Novartis, Pfizer and Sunovion.

## REFERENCES

1. Rabe, K.F., Hurd, S., Anzueto, A., Barnes, P.J., Buist, S.A., Calverley, P., Fukuchi, Y., Jenkins, C., Rodriguez-Roisin, R., van Weel, C. *et al.* (2007) Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am. J. Respir. Crit. Care Med.*, **176**, 532–555.
2. Miniño, A.M., Xu, J. and Kochanek, K.D. (2010) *National Vital Statistics Reports: Deaths: Preliminary Data for 2008*. National Center for Health Statistics, Hyattsville, Maryland, Vol. 59.
3. Lokke, A., Lange, P., Scharling, H., Fabricius, P. and Vestbo, J. (2006) Developing COPD: a 25 year follow up study of the general population. *Thorax*, **61**, 935–939.
4. Burrows, B., Knudson, R.J., Cline, M.G. and Lebowitz, M.D. (1977) Quantitative relationships between cigarette smoking and ventilatory function. *Am. Rev. Respir. Dis.*, **115**, 195–205.
5. Silverman, E.K., Chapman, H.A., Drazen, J.M., Weiss, S.T., Rosner, B., Campbell, E.J., O'Donnell, W.J., Reilly, J.J., Ginns, L., Mentzer, S. *et al.* (1998) Genetic epidemiology of severe, early-onset chronic obstructive pulmonary disease. Risk to relatives for airflow obstruction and chronic bronchitis. *Am. J. Respir. Crit. Care Med.*, **157**, 1770–1778.
6. McCloskey, S.C., Patel, B.D., Hinchliffe, S.J., Reid, E.D., Wareham, N.J. and Lomas, D.A. (2001) Siblings of patients with severe chronic obstructive pulmonary disease have a significant risk of airflow obstruction. *Am. J. Respir. Crit. Care Med.*, **164**, 1419–1424.
7. Palmer, L.J., Celedon, J.C., Chapman, H.A., Speizer, F.E., Weiss, S.T. and Silverman, E.K. (2003) Genome-wide linkage analysis of bronchodilator responsiveness and post-bronchodilator spirometric phenotypes in chronic obstructive pulmonary disease. *Hum. Mol. Genet.*, **12**, 1199–1210.
8. Silverman, E.K., Mosley, J.D., Palmer, L.J., Barth, M., Senter, J.M., Brown, A., Drazen, J.M., Kwiatkowski, D.J., Chapman, H.A., Campbell, E.J. *et al.* (2002) Genome-wide linkage analysis of severe, early-onset chronic obstructive pulmonary disease: airflow obstruction and chronic bronchitis phenotypes. *Hum. Mol. Genet.*, **11**, 623–632.
9. Castaldi, P.J., Cho, M.H., Cohn, M., Langerman, F., Moran, S., Tarragona, N., Moukhachen, H., Venugopal, R., Hasimja, D., Kao, E.

- et al.* (2010) The COPD genetic association compendium: a comprehensive online database of COPD genetic associations. *Hum. Mol. Genet.*, **19**, 526–534.
10. Hersh, C.P., Demeo, D.L., Lange, C., Litonjua, A.A., Reilly, J.J., Kwiatkowski, D., Laird, N., Sylvia, J.S., Sparrow, D., Speizer, F.E. *et al.* (2005) Attempted replication of reported chronic obstructive pulmonary disease candidate gene associations. *Am. J. Respir. Cell Mol. Biol.*, **33**, 71–78.
  11. Pillai, S.G., Ge, D., Zhu, G., Kong, X., Shianna, K.V., Need, A.C., Feng, S., Hersh, C.P., Bakke, P., Gulsvik, A. *et al.* (2009) A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet.*, **5**, e1000421.
  12. Wilk, J.B., Chen, T.H., Gottlieb, D.J., Walter, R.E., Nagle, M.W., Brandler, B.J., Myers, R.H., Borecki, I.B., Silverman, E.K., Weiss, S.T. *et al.* (2009) A genome-wide association study of pulmonary function measures in the Framingham Heart Study. *PLoS Genet.*, **5**, e1000429.
  13. Cho, M.H., Boutaoui, N., Klanderman, B.J., Sylvia, J.S., Ziniti, J.P., Hersh, C.P., DeMeo, D.L., Hunninghake, G.M., Litonjua, A.A., Sparrow, D. *et al.* (2010) Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat. Genet.*, **42**, 200–202.
  14. Hancock, D.B., Eijgelsheim, M., Wilk, J.B., Gharib, S.A., Loehr, L.R., Marcante, K.D., Franceschini, N., van Durme, Y.M., Chen, T.H., Barr, R.G. *et al.* (2010) Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat. Genet.*, **42**, 45–52.
  15. Repapi, E., Sayers, I., Wain, L.V., Burton, P.R., Johnson, T., Obeidat, M., Zhao, J.H., Ramasamy, A., Zhai, G., Vitart, V. *et al.* (2010) Genome-wide association study identifies five loci associated with lung function. *Nat. Genet.*, **42**, 36–44.
  16. Lambrechts, D., Buyschaert, I., Zanen, P., Coolen, J., Lays, N., Cuppens, H., Groen, H.J., Dewever, W., van Klaveren, R.J., Verschakelen, J. *et al.* (2010) The 15q24/25 susceptibility variant for lung cancer and chronic obstructive pulmonary disease is associated with emphysema. *Am. J. Respir. Crit. Care Med.*, **181**, 486–493.
  17. Young, R.P., Hopkins, R.J., Whittington, C.F., Hay, B.A., Epton, M.J. and Gamble, G.D. (2011) Individual and cumulative effects of GWAS susceptibility loci in lung cancer: associations after sub-phenotyping for COPD. *PLoS ONE*, **6**, e16476.
  18. Van Durme, Y.M., Eijgelsheim, M., Joos, G.F., Hofman, A., Uitterlinden, A.G., Brusselle, G.G. and Stricker, B.H. (2010) Hedgehog-interacting protein is a COPD susceptibility gene: the Rotterdam Study. *Eur. Respir. J.*, **36**, 89–95.
  19. Young, R.P., Hopkins, R.J., Hay, B.A., Epton, M.J., Black, P.N. and Gamble, G.D. (2008) Lung cancer gene associated with COPD: triple whammy or possible confounding effect? *Eur. Respir. J.*, **32**, 1158–1164.
  20. Young, R.P., Whittington, C.F., Hopkins, R.J., Hay, B.A., Epton, M.J., Black, P.N. and Gamble, G.D. (2010) Chromosome 4q31 locus in COPD is also associated with lung cancer. *Eur. Respir. J.*, **36**, 1375–1382.
  21. Chappell, S.L., Daly, L., Lotya, J., Alsaegh, A., Guetta-Baranes, T., Roca, J., Rabinovich, R., Morgan, K., Millar, A.B., Donnelly, S.C. *et al.* (2011) The role of IREB2 and transforming growth factor beta-1 genetic variants in COPD: a replication case-control study. *BMC Med. Genet.*, **12**, 24.
  22. Han, B. and Eskin, E. (2011) Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am. J. Hum. Genet.*, **88**, 586–598.
  23. Thorgeirsson, T.E., Gudbjartsson, D.F., Surakka, I., Vink, J.M., Amin, N., Geller, F., Sulem, P., Rafnar, T., Esko, T., Walter, S. *et al.* (2010) Sequence variants at CHRN3-CHRNA6 and CYP2A6 affect smoking behavior. *Nat. Genet.*, **42**, 448–453.
  24. Tobacco and Genetics (TAG) Consortium. (2010) Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat. Genet.*, **42**, 441–447.
  25. Saccone, N.L., Culverhouse, R.C., Schwantes-An, T.H., Cannon, D.S., Chen, X., Cichon, S., Giegling, I., Han, S., Han, Y., Keskitalo-Vuokko, K. *et al.* (2010) Multiple independent loci at chromosome 15q25.1 affect smoking quantity: a meta-analysis and comparison with lung cancer and COPD. *PLoS Genet.*, **6**, pii: e1001053.
  26. Siedlinski, M., Cho, M.H., Bakke, P., Gulsvik, A., Lomas, D.A., Anderson, W., Kong, X., Rennard, S.I., Beaty, T.H., Hokanson, J.E. *et al.* (2011) Genome-wide association study of smoking behaviours in patients with COPD. *Thorax*, **66**, 894–902.
  27. Larson, R.K. and Barman, M.L. (1965) The familial occurrence of chronic obstructive pulmonary disease. *Ann. Intern. Med.*, **63**, 1001–1008.
  28. Laurell, C.-B. and Eriksson, A. (1963) The electrophoretic alpha-1 globulin pattern of serum in alpha 1-antitrypsin deficiency. *Scand. J. Clin. Lab. Invest.*, **15**, 132–140.
  29. Wang, J., Spitz, M.R., Amos, C.I., Wilkinson, A.V., Wu, X. and Shete, S. (2010) Mediating effects of smoking and chronic obstructive pulmonary disease on the relation between the CHRNA5-A3 genetic locus and lung cancer risk. *Cancer*, **116**, 3458–3462.
  30. DeMeo, D.L., Mariani, T., Bhattacharya, S., Srisuma, S., Lange, C., Litonjua, A., Bueno, R., Pillai, S.G., Lomas, D.A., Sparrow, D. *et al.* (2009) Integration of genomic and genetic approaches implicates IREB2 as a COPD susceptibility gene. *Am. J. Hum. Genet.*, **85**, 493–502.
  31. Otulakowski, G., Duan, W. and O'Brodovich, H. (2009) Global and gene-specific translational regulation in rat lung development. *Am. J. Respir. Cell Mol. Biol.*, **40**, 555–567.
  32. Lin, S., Ikegami, M., Xu, Y., Bosserhoff, A.K., Malkinson, A.M. and Shannon, J.M. (2008) Misexpression of MIA disrupts lung morphogenesis and causes neonatal death. *Dev. Biol.*, **316**, 441–455.
  33. Groenman, F., Rutter, M., Caniggia, I., Tibboel, D. and Post, M. (2007) Hypoxia-inducible factors in the first trimester human lung. *J. Histochem. Cytochem.*, **55**, 355–363.
  34. Hersh, C.P., Demeo, D.L., Lazarus, R., Celedon, J.C., Raby, B.A., Benditt, J.O., Criner, G., Make, B., Martinez, F.J., Scanlon, P.D. *et al.* (2006) Genetic association analysis of functional impairment in chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.*, **173**, 977–984.
  35. Sterner-Kock, A., Thorey, I.S., Koli, K., Wempe, F., Otte, J., Bangsow, T., Kuhlmeier, K., Kirchner, T., Jin, S., Keski-Oja, J. *et al.* (2002) Disruption of the gene encoding the latent transforming growth factor-beta binding protein 4 (LTBP-4) causes abnormal lung development, cardiomyopathy, and colorectal cancer. *Genes Dev.*, **16**, 2264–2273.
  36. Urban, Z., Huchtagowder, V., Schurmann, N., Todorovic, V., Zilberberg, L., Choi, J., Sens, C., Brown, C.W., Clark, R.D., Holland, K.E. *et al.* (2009) Mutations in LTBP4 cause a syndrome of impaired pulmonary, gastrointestinal, genitourinary, musculoskeletal, and dermal development. *Am. J. Hum. Genet.*, **85**, 593–605.
  37. London, S.J., Idle, J.R., Daly, A.K. and Coetzee, G.A. (1999) Genetic variation of CYP2A6, smoking, and risk of cancer. *Lancet*, **353**, 898–899.
  38. Nakajima, M., Yamamoto, T., Nunoya, K., Yokoi, T., Nagashima, K., Inoue, K., Funae, Y., Shimada, N., Kamataki, T. and Kuroiwa, Y. (1996) Characterization of CYP2A6 involved in 3'-hydroxylation of cotinine in human liver microsomes. *J. Pharmacol. Exp. Ther.*, **277**, 1010–1015.
  39. Hukkanen, J., Jacob, P. 3rd and Benowitz, N.L. (2005) Metabolism and disposition kinetics of nicotine. *Pharmacol. Rev.*, **57**, 79–115.
  40. Al Koudsi, N., Ahluwalia, J.S., Lin, S.K., Sellers, E.M. and Tyndale, R.F. (2009) A novel CYP2A6 allele (CYP2A6\*35) resulting in an amino-acid substitution (Asn438Tyr) is associated with lower CYP2A6 activity *in vivo*. *Pharmacogenomics J.*, **9**, 274–282.
  41. Mwenifumbo, J.C., Al Koudsi, N., Ho, M.K., Zhou, Q., Hoffmann, E.B., Sellers, E.M. and Tyndale, R.F. (2008) Novel and established CYP2A6 alleles impair *in vivo* nicotine metabolism in a population of Black African descent. *Hum. Mutat.*, **29**, 679–688.
  42. Bloom, J., Hinrichs, A.L., Wang, J.C., von Weyarn, L.B., Kharasch, E.D., Bierut, L.J., Goate, A. and Murphy, S.E. (2011) The contribution of common CYP2A6 alleles to variation in nicotine metabolism among European-Americans. *Pharmacogenet. Genomics*, **21**, 403–416.
  43. Rao, Y., Hoffmann, E., Zia, M., Bodin, L., Zeman, M., Sellers, E.M. and Tyndale, R.F. (2000) Duplications and defects in the CYP2A6 gene: identification, genotyping, and *in vivo* effects on smoking. *Mol. Pharmacol.*, **58**, 747–755.
  44. Le Marchand, L., Derby, K.S., Murphy, S.E., Hecht, S.S., Hatsukami, D., Carmella, S.G., Tiirikainen, M. and Wang, H. (2008) Smokers with the CHRNA lung cancer-associated variants are exposed to higher levels of nicotine equivalents and a carcinogenic tobacco-specific nitrosamine. *Cancer Res.*, **68**, 9137–9140.
  45. Timofeeva, M.N., McKay, J.D., Smith, G.D., Johansson, M., Byrnes, G.B., Chabrier, A., Relton, C., Ueland, P.M., Vollset, S.E., Midttun, O. *et al.* (2011) Genetic polymorphisms in 15q25 and 19q13 loci, cotinine levels, and risk of lung cancer in EPIC. *Cancer Epidemiol. Biomarkers Prev.*, **20**, 2250–2261.
  46. Bornaards, C.M., Twisk, J.W., Snel, J., Van Mechelen, W. and Kemper, H.C. (2001) Is calculating pack-years retrospectively a valid method to estimate life-time tobacco smoking? A comparison between prospectively calculated pack-years and retrospectively calculated pack-years. *Addiction*, **96**, 1653–1661.

47. Njolstad, I., Arnesen, E. and Lund-Larsen, P.G. (1996) Smoking, serum lipids, blood pressure, and sex differences in myocardial infarction. A 12-year follow-up of the Finnmark Study. *Circulation*, **93**, 450–456.
48. Schunkert, H., König, I.R., Kathiresan, S., Reilly, M.P., Assimes, T.L., Holm, H., Preuss, M., Stewart, A.F., Barbalic, M., Gieger, C. *et al.* (2011) Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.*, **43**, 333–338.
49. Kohansal, R., Martinez-Camblor, P., Agusti, A., Buist, A.S., Mannino, D.M. and Soriano, J.B. (2009) The natural history of chronic airflow obstruction revisited: an analysis of the Framingham offspring cohort. *Am. J. Respir. Crit. Care Med.*, **180**, 3–10.
50. Kanner, R.E., Connett, J.E., Williams, D.E. and Buist, A.S. (1999) Effects of randomized assignment to a smoking cessation intervention and changes in smoking habits on respiratory symptoms in smokers with early chronic obstructive pulmonary disease: the Lung Health Study. *Am. J. Med.*, **106**, 410–416.
51. Anthonisen, N.R., Skeans, M.A., Wise, R.A., Manfreda, J., Kanner, R.E. and Connett, J.E. (2005) The effects of a smoking cessation intervention on 14.5-year mortality: a randomized clinical trial. *Ann. Intern. Med.*, **142**, 233–239.
52. Anthonisen, N.R., Connett, J.E., Kiley, J.P., Altose, M.D., Bailey, W.C., Buist, A.S., Conway, W.A. Jr, Enright, P.L., Kanner, R.E., O'Hara, P. *et al.* (1994) Effects of smoking intervention and the use of an inhaled anticholinergic bronchodilator on the rate of decline of FEV1. The Lung Health Study. *JAMA*, **272**, 1497–1505.
53. Carpenter, M.J., Strange, C., Jones, Y., Dickson, M.R., Carter, C., Moseley, M.A. and Gilbert, G.E. (2007) Does genetic testing result in behavioral health change? Changes in smoking behavior following testing for alpha-1 antitrypsin deficiency. *Ann. Behav. Med.*, **33**, 22–28.
54. Audrain, J., Boyd, N.R., Roth, J., Main, D., Caporaso, N.F. and Lerman, C. (1997) Genetic susceptibility testing in smoking-cessation treatment: one-year outcomes of a randomized trial. *Addict. Behav.*, **22**, 741–751.
55. Risch, N. and Merikangas, K. (1996) The future of genetic studies of complex human diseases. *Science*, **273**, 1516–1517.
56. Evangelou, E., Trikalinos, T.A., Salanti, G. and Ioannidis, J.P. (2006) Family-based versus unrelated case-control designs for genetic associations. *PLoS Genet.*, **2**, e123.
57. Kelleher, C.M., Silverman, E.K., Broekelmann, T., Litonjua, A.A., Hernandez, M., Sylvia, J.S., Stoler, J., Reilly, J.J., Chapman, H.A., Speizer, F.E. *et al.* (2005) A functional mutation in the terminal exon of elastin in severe, early-onset chronic obstructive pulmonary disease. *Am. J. Respir. Cell Mol. Biol.*, **33**, 355–362.
58. Patel, B.D., Coxson, H.O., Pillai, S.G., Agusti, A.G., Calverley, P.M., Donner, C.F., Make, B.J., Muller, N.L., Rennard, S.I., Vestbo, J. *et al.* (2008) Airway wall thickening and emphysema show independent familial aggregation in chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.*, **178**, 500–505.
59. Silverman, E.K., Vestbo, J., Agusti, A., Anderson, W., Bakke, P.S., Barnes, K.C., Graham Barr, R., Bleecker, E.R., Boezen, H.M., Burkart, K.M. *et al.* (2011) Opportunities and challenges in the genetics of COPD 2010: an international COPD genetics conference report. *COPD*, **8**, 121–135.
60. Sorheim, I.C., Johannessen, A., Grydeland, T.B., Omenaas, E.R., Gulsvik, A. and Bakke, P.S. (2010) Case-control studies on risk factors for chronic obstructive pulmonary disease: how does the sampling of the cases and controls affect the results? *Clin. Respir. J.*, **4**, 89–96.
61. Vestbo, J., Anderson, W., Coxson, H.O., Crim, C., Dawber, F., Edwards, L., Hagan, G., Knobil, K., Lomas, D.A., MacNee, W. *et al.* (2008) Evaluation of COPD longitudinally to identify predictive surrogate end-points (ECLIPSE). *Eur. Respir. J.*, **31**, 869–873.
62. Bell, B., Rose, C. and Damon, H. (1972) The Normative Aging Study: an interdisciplinary and longitudinal study of health and aging. *Aging Hum. Dev.*, **3**, 5–17.
63. Fishman, A., Martinez, F., Naunheim, K., Piantadosi, S., Wise, R., Ries, A., Weinmann, G. and Wood, D.E. (2003) A randomized trial comparing lung-volume-reduction surgery with medical therapy for severe emphysema. *N. Engl. J. Med.*, **348**, 2059–2073.
64. Zhu, G., Warren, L., Aponte, J., Gulsvik, A., Bakke, P., Anderson, W.H., Lomas, D.A., Silverman, E.K. and Pillai, S.G. (2007) The SERPINE2 gene is associated with chronic obstructive pulmonary disease in two large populations. *Am. J. Respir. Crit. Care Med.*, **176**, 167–173.
65. Regan, E.A., Hokanson, J.E., Murphy, J.R., Make, B., Lynch, D.A., Beaty, T.H., Curran-Everett, D., Silverman, E.K. and Crapo, J.D. (2010) Genetic epidemiology of COPD (COPDGene) study design. *COPD*, **7**, 32–43.
66. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.*, **38**, 904–909.
67. Patterson, N., Price, A.L. and Reich, D. (2006) Population structure and eigenanalysis. *PLoS Genet.*, **2**, e190.
68. Li, Y., Willer, C.J., Ding, J., Scheet, P. and Abecasis, G.R. (2010) MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.*, **34**, 816–834.
69. Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M. *et al.* (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature*, **449**, 851–861.
70. 1000 Genomes Project Consortium. (2010) A map of human genome variation from population-scale sequencing. *Nature*, **467**, 1061–1073.
71. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
72. de Bakker, P.I., Ferreira, M.A., Jia, X., Neale, B.M., Raychaudhuri, S. and Voight, B.F. (2008) Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum. Mol. Genet.*, **17**, R122–R128.
73. Willer, C.J., Li, Y. and Abecasis, G.R. (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, **26**, 2190–2191.
74. Devlin, B. and Roeder, K. (1999) Genomic control for association studies. *Biometrics*, **55**, 997–1004.
75. Aulchenko, Y.S., Ripke, S., Isaacs, A. and van Duijn, C.M. (2007) GenABEL: an R library for genome-wide association analysis. *Bioinformatics*, **23**, 1294–1296.
76. Lange, C., DeMeo, D., Silverman, E.K., Weiss, S.T. and Laird, N.M. (2004) PBAT: tools for family-based association studies. *Am. J. Hum. Genet.*, **74**, 367–369.
77. Lin, S., Chakravarti, A. and Cutler, D.J. (2004) Exhaustive allelic transmission disequilibrium tests as a new approach to genome-wide association studies. *Nat. Genet.*, **36**, 1181–1188.
78. Voight, B.F., Scott, L.J., Steinthorsdottir, V., Morris, A.P., Dina, C., Welch, R.P., Zeggini, E., Huth, C., Aulchenko, Y.S., Thorleifsson, G. *et al.* (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.*, **42**, 579–589.
79. Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F. and Sklar, P. (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, **460**, 748–752.
80. So, H.C., Gui, A.H., Cherny, S.S. and Sham, P.C. (2011) Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. *Genet. Epidemiol.*, **35**, 310–317.
81. Buist, A.S., McBurnie, M.A., Vollmer, W.M., Gillespie, S., Burney, P., Mannino, D.M., Menezes, A.M., Sullivan, S.D., Lee, T.A., Weiss, K.B. *et al.* (2007) International variation in the prevalence of COPD (the BOLD Study): a population-based prevalence study. *Lancet*, **370**, 741–750.
82. Johnson, A.D., Handsaker, R.E., Pulit, S.L., Nizzari, M.M., O'Donnell, C.J. and de Bakker, P.I. (2008) SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*, **24**, 2938–2939.
83. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Glied, T.P., Boehnke, M., Abecasis, G.R. and Willer, C.J. (2010) LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*, **26**, 2336–2337.