CONFERENCE REPORT

Opportunities and Challenges in the Genetics of COPD 2010: An International COPD Genetics Conference Report


Channing Laboratory (EKS, DLD, CPH, AAL) and Pulmonary and Critical Care Division (EKS, BRC, DLD, CPH, AAL), Brigham and Women’s Hospital, Harvard Medical School, Boston, Massachusetts, USA,1 University of Copenhagen / Hvidovre Hospital, Hvidovre, Denmark,2 Institut del Tòrax Hospital Clinic, Universitat de Barcelona, and Centro de Investigación Biomédica en Red de Enfermedades Respiratorias (CIBERES), Barcelona, Spain,3 GlaxoSmithKline, Research Triangle Park, NC,4 Institute of Internal Medicine, University of Bergen, Bergen, Norway,5 Johns Hopkins Asthma & Allergy Center (KCB and NNH) and Department of Epidemiology (MAP), Johns Hopkins University, Baltimore, Maryland, USA,6 Department of Medicine (RGB and KMB) and Department of Epidemiology (RGB), Columbia University, Columbia University Medical Center, New York, New York, USA,7 Wake Forest University, Winston-Salem, North Carolina, USA,8 Department of Pulmonology (DSP), University Medical Center Groningen, Groningen, the Netherlands,9 Division of Respiratory Sciences, Imperial College, London, United Kingdom,10 Division of Lung Diseases, National Heart, Lung, and Blood Institute, Bethesda, Maryland, USA,11 University of British Columbia, James Hogg iCAPTURE Center for Cardiovascular and Pulmonary Research, Vancouver, British Columbia, Canada,12 Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain; Municipal Institute of Medical Research (IMIM-Hospital del Mar), Barcelona, Spain; Department of Experimental and Health Sciences, Universitat Pompeu Fabra, Barcelona, Spain; CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain,13 Division of Therapeutics and Molecular Medicine (IPH) and Molecular Medical Sciences (NK), Nottingham Respiratory Biomedical Research Unit, University Hospital of Nottingham, Nottingham, United Kingdom,14 Kangwon National University, Chunchun, South Korea,15 Vesalius Research Center, VIB and University of Leuven, Leuven, Belgium,16 Center for Chronic Obstructive Airways Disease, Division of Pulmonary and Critical Care Medicine, Asan Medical Center, Seoul, South Korea,17 Department of Medicine, University of Cambridge, Cambridge Institute for Medical Research, Cambridge, United Kingdom,18 Division of Intramural Research, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Research Triangle Park, North Carolina, USA,19 Hokkaido University School of Medicine, Sapporo, Japan,20 University of Copenhagen and Dept. of Clinical Biochemistry, Herlev Hospital, Copenhagen University Hospital, Herlev, Denmark,21 Division of Intramural Research, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, MD, and NHLBI’s Framingham Heart Study, Framingham, Massachusetts, USA,22 Lovelace Respiratory Research Institute, Albuquerque, New Mexico, USA,23 Departments of Health Sciences & Genetics, University of Leicester, United Kingdom,24 Division for Pulmonary Diseases, University of Marburg, Marburg, Germany,25 Pulmonary Center, Boston University School of Medicine, Boston, Massachusetts, USA,26 Department of Respiratory Diseases, University Hospital Maastricht, Maastricht, the Netherlands,27 University of Auckland and Synergy Bioscience Ltd, Auckland, New Zealand,28 Helmholtz Zentrum München, Germany,29 University of Edinburgh, MRC Centre for Inflammation Research, The Queen’s Medical Research Institute, Edinburgh, United Kingdom,30 National Jewish Health, Denver, Colorado, USA31

Keywords: COPD, Genetics, Association analysis, Consortium

Conference sponsored by the National Heart, Lung and Blood Institute and the COPD Foundation.

Correspondence to: Edwin K. Silverman, Channing Laboratory, Brigham and Women’s Hospital, 181 Longwood Avenue, Boston, MA 02115, phone: 617–525–0856, fax: 617–525–0958. email: ed.silverman@channing.harvard.edu
CONFERENCE PARTICIPANTS

Alvar Agusti, MD, PhD
Wayne Anderson, PhD
Per S. Bakke, PhD, MD
Kathleen C. Barnes, PhD
R. Graham Barr, MD, DrPH
Eugene R. Bleeker, MD
Kristin M. Burkart, MD, MSc
Bartolome R. Celi, MD
Michael H. Cho, MD
William O.C. Cookson, MD, DPhil
James D. Crapo, MD
Thomas Croxton, MD, PhD
Denise Daley, PhD
Dawn L. DeMeo, MD, MPH
Weiniu Gan, PhD
Judith Garcia-Aymerich, MD
Nadia N. Hansel, MD, MDPh
Craig Hersh, MD, MPH
Noor Kalshkeker, MD
James P. Kiley, PhD
Woo Jin Kim, MD, PhD
Diether Lambrechts, MD
Sang-Do Lee, MD
Stephanie J. London, M.D., Dr.P.H.
William MacNee, MB, ChB, MD
Masaharu Nishimura, MD
Borge G. Nordestgaard, MD, DMSc
Christopher O’Donnell, M.D.
Edwin K. Silverman, MD, PhD
Yohannes Tesfaigzi, PhD
Martin D. Tobin, MD
Jørgen Vestbo, MD, DMSc
Claus Vogelmeier, MD
John W. Walsh
Jemma B. Wilk, DSc
Emiel Wouters, MD, PhD
Robert P. Young, MD, PhD
Loems Ziegler-Heitbrock, MD

REPORT AND RECOMMENDATIONS FROM CONFERENCE
HELD JULY 13–14, 2010, BOSTON, MASSACHUSETTS

INTRODUCTION

Chronic obstructive pulmonary disease (COPD) is defined by the Global Initiative for Chronic Obstructive Lung Disease (GOLD) as a disease state characterized by airflow limitation that is not fully reversible (1). Cigarette smoking is the most important risk factor for the development of COPD. Although the dose-response relationship between cigarette smoking and pulmonary function is well-established, there is considerable variability in the reduction in FEV1 among smokers with similar smoking exposures (2, 3). The low percentage of variance in pulmonary function explained by smoking suggests that there could be genetic differences in susceptibility to the effects of cigarette smoking (4, 5). In addition to genetic factors, other environmental determinants such as indoor biomass smoke exposure can be important risk factors for COPD (6). A small percentage of COPD patients (estimated at 1–2%) inherit severe alpha-1 antitrypsin (AAT) deficiency, which proves that genetic factors can influence COPD susceptibility. The discovery of AAT deficiency was a major factor in the development of the Protease-Antiprotease Hypothesis for COPD, which has been one of the prevailing models of disease pathogenesis for more than 40 years.

With the substantial impact of AAT deficiency on our understanding of COPD pathogenesis, it was natural to hope that the identification of other COPD susceptibility genes would lead to similar novel insights into COPD. Until recently, however, progress in the identification of additional genetic risk factors for COPD has been slow.

To facilitate the development of such research, a meeting of COPD genetics investigators was held on July 13–14, 2010 in Boston. The goals of the meeting were:

1. To review the current state of COPD genetics research;
2. To discuss existing study populations for COPD genetics research throughout the world;
3. To consider opportunities for collaborations between different COPD research groups through an International COPD Genetics Consortium;
4. To recognize challenges in building COPD genetics collaborations and to discuss them openly; and,
5. To develop a framework for future collaborative studies.

Current status of COPD genetics research

Many candidate gene association studies have been performed over the past 40 years, but the results have been largely inconsistent. These inconsistencies likely relate to a variety of methodological issues, including small sample sizes, variable definitions of case and control groups, failure to adjust for multiple statistical testing, and inadequate adjustments for population stratification and smoking exposure. Most of the studies describing COPD-associated polymorphisms were performed in White populations (7). A meta-analysis of 20 polymorphisms in 12 candidate genes involved in the protease–antiprotease balance and several antioxidant pathways showed that, after combining independent studies, many of these candidate genes had no association with COPD (8).

Another factor likely impeding the progress of identifying COPD susceptibility genes is the lack of accurate phenotypic characterization of this complex and heterogeneous disease. Airflow limitation determined by spirometry has been the most common approach to classify and monitor the disease. Structural changes of the lung including emphysema and small airway obstruction are the primary processes that affect lung function (9), but they are not easily discernable with the simple spirometric measures commonly used for phenotyping COPD. Recent advances in characterizing pathologic changes such as emphysema and remodeling of the small and large airways by quantitative analyses of image data from multidetector computed tomography (CT), together with physiological testing, have been helpful to differentiate COPD phenotypes (emphysema-predominant, airway-predominant, or mixed) (10). Study populations that have chest CT data may help to better identify COPD-associated genetic variations (11). Other potentially relevant COPD
phenotypes, such as cachexia and low exercise capacity, have not been widely analyzed in COPD genetic studies.

Perhaps the greatest problem in the candidate gene era of COPD genetic studies was improper candidate gene selection, which reflects our limited understanding of COPD pathogenesis. However, the application of genome-wide association studies (GWAS), which provide an unbiased and comprehensive search throughout the genome for common susceptibility loci, has changed the landscape of COPD genetics. Based on GWAS, three genetic loci have been unequivocally associated with COPD susceptibility, located on chromosome 4 near the HHIP gene, on chromosome 4 in the FAM13A gene, and on chromosome 15 in a block of genes which contains several components of the nicotinic acetylcholine receptor as well as the IREB2 gene.

In 2009, a series of studies provided convincing support for these three genetic loci in COPD susceptibility. Pillai and colleagues found genome-wide significant associations of the CHRNA3/CHRNAS/IREB2 region to COPD (12). DeMeo and colleagues performed gene expression studies of normal vs. COPD lung tissues followed by genetic association analysis of COPD (13), suggesting that at least one of the key COPD genetic determinants in the chromosome 15 GWAS region was IREB2.

In the Framingham Heart Study (14), the HHIP region was associated with FEV1/FVC at genome-wide significance with replication of the effect on FEV1/FVC demonstrated in an independent sample drawn from the Family Heart Study, and this same region nearly reached genome-wide significance with COPD susceptibility in the Pillai paper (12). Recently, two papers published in Nature Genetics from large general population samples have provided strong support for the association of HHIP SNPs with FEV1/FVC (15, 16). One of these articles, from the CHARGE Consortium, also found evidence for association of FEV1/FVC with the FAM13A locus (15), which has been strongly associated with COPD susceptibility (17).

Moreover, several case-control studies from other European populations have replicated these findings by confirming significant associations to the chromosome 15q25 locus (CHRNA3/CHRNAS/IREB2) (18, 19), chromosome 4q31 locus (HHIP) (20, 21), and chromosome 4q22 locus (FAM13A) (22). Thus, the frustration of inconsistent genetic association results in COPD from the beginning of the last decade has been replaced by optimism regarding the likely importance of the IREB2/CHRNA3/CHRNAS, HHIP, and FAM13A loci in COPD susceptibility.

Advantages of creating large networks for genetic analysis

There are likely multiple additional COPD susceptibility genetic determinants that have not yet been identified. In many other complex diseases, the creation of large collaborative consortia has enabled highly powered genome-wide association studies that have led to the identification of multiple novel genetic susceptibility loci. For example, a Type 2 Diabetes mellitus consortium performed GWAS in 8,130 Cases and 38,987 Controls and identified multiple novel susceptibility loci (23). The International Lung Cancer Consortium found new SNPs that were associated with disease in Asian populations (24, 25). The ENGAGE consortium discovered sequence variants associated with smoking behavior within regions harboring nAChR genes (CHRN3–CHRNAS, 8p11) and a nicotine-metabolizing enzyme (26). We anticipate that a similar collaborative consortium approach in COPD could lead to the identification of additional novel COPD genetic determinants.

Gaps in current genetic knowledge

The most fundamental gap in current COPD genetics knowledge is that there are probably many genetic determinants of COPD, but only three genomic regions likely to contain such susceptibility loci have been conclusively identified. Moreover, the functional genetic variants within the three existing COPD GWAS regions remain to be found. To adequately analyze the various subtypes of COPD, studies that include multiple ethnic groups as well as multiple environmental factors that influence inflammation will be required in large sample sizes. More recently, some studies have combined results from several populations to increase the numbers of cases and controls. In more than 8300 subjects in seven study populations, the minor allele of a SNP in MMP12 was associated with a positive effect on lung function and a reduced risk of COPD (27). The genome-wide association study that identified FAM13A included three sets of COPD cases and smoking controls (17). However, these studies are still underpowered to identify genetic determinants of small effect, and establishing a consortium of groups studying cigarette smokers may facilitate pooling large samples to identify genetic variants associated with COPD susceptibility.

GENETIC TECHNOLOGIES AVAILABLE FOR AN INTERNATIONAL COPD GENETICS CONSORTIUM

It is desirable that the full power of modern genetic and genomic technology and techniques be brought to bear on COPD. Statistical genetic approaches should begin with meta-analyses of currently completed GWA studies, including imputation of polymorphisms from the 1000 Genomes Project. Analyses should routinely include epidemiologically important covariates such as sex, age at onset, and smoking history. Ancestry needs to be matched carefully between cases and controls, using, for example, principal component analyses. Multi-marker techniques to identify polygenic effects below the GWAS threshold may be useful in identifying genes and pathways impacting on the disease.

Genome-wide SNP genotyping of several thousand or more cases is necessary, particularly using existing European panels of subjects that have not yet been genotyped and cases and controls of non-European ancestry. It is noted that there exists a wide range of previously genotyped European controls that could be used wherever possible.

Further meta-analysis of the full dataset should be completed after the additional genotyping. Ideally these results would be integrated with large-scale studies of other smoking-related diseases (particularly lung cancer and cardiovascular disease), with studies of smoking behavior and
addiction, and studies of diseases characterized by compromised lung function (in particular, asthma).

Fine mapping of selected loci to identify functional variants will be necessary. This will include statistical approaches such as multiple regression as well as additional genotyping. The particular importance of including ancestral groups of non-European origin in these analyses is noted, in order to use their differences in linkage disequilibrium patterns to break up linkage disequilibrium blocks, and to demonstrate generalizability of variants associated with COPD to the population at large.

Next-generation DNA sequencing approaches have the capacity to discover highly penetrant rare variants in common diseases such as COPD. Limiting sequencing to the exome greatly reduces costs compared to whole genome sequencing approaches, while retaining much of the information that is likely to lead to the identification of disease-related rare variants. Although the value of exome sequencing has not yet been established in complex genetic diseases, it is desirable to explore the use of exome sequencing to search for rare mutations in patients with severe spectrum disease, including non-smokers with COPD as a separate group.

Genomic studies allow systematic investigation of pathways and networks of gene functions underlying disease (28). Investigations for COPD should include mapping of expression quantitative trait loci (eQTL) and network identification from measurements of global gene expression in airway biopsies and peripheral blood DNA samples. It would also be important to carry out eQTL mapping and network identification with global gene expression in current cigarette smokers and non-smokers. The investigation of methylQTL (using genome-wide methylation arrays) should similarly be implemented in order to explore epigenetic effects on COPD and related phenotypes.

Lastly, it is now possible to quantify bacterial colonization of airways using DNA and RNA sequencing techniques that address the hyper-variable bacterial 16S gene as well as metagenomic approaches that examine the global gene content and gene expression of human bacteria (29). It is therefore recommended that systematic studies of the microbiome be carried out in patients with COPD. These studies should include 16S sequencing for bacterial identification; metagenomic sequencing and measures of bacterial gene expression; and investigation of relationships of these measures to host gene expression and genotype.

COPD PHENOTYPING AND KNOWLEDGE GAPS

Clinical phenotypes

Precise definition and validation of clinical phenotypes are key prerequisites to identify the genetic basis of complex diseases, since a principal goal of genetic research is to identify specific genotypes that link to specific phenotypes (30, 31). From the genetics point of view, if current approaches in defining phenotypes are inadequate, the huge amount of currently available genotypic data cannot be optimally used (32). A recent consensus definition (11) proposes that a “COPD clinical phenotype” is “a single or combination of disease attributes that describe differences between individuals with COPD as they relate to clinically meaningful outcomes (symptoms, exacerbations, response to therapy, rate of disease progression or death).” Thus, for a COPD phenotype to be of use in a COPD genetics study, it has to be associated with clinically meaningful outcomes. Some inconsistent results published so far on the genetic basis of COPD may be due to the lack of an appropriate characterization of different clinical COPD phenotypes (intra-study variation), as well as to ethnic differences among studies (inter-study variation) (33).

The degree of airflow limitation remains the defining characteristic of COPD and thus its most important phenotypic expression. However, there is sufficient evidence to support the need to consider additional phenotypic expressions in the characterization of patients with COPD. These include: 1) the degree, type and distribution of emphysema (discussed below); 2) the extent of airway wall thickening caused by inflammation; 3) the degree of hyperinflation expressed by the IC and the IC/TLC; 4) the presence of abnormal gas exchange (hypoxia and hypercapnia); 5) the presence of systemic involvement as measured by the BMI; 6) the exercise capacity whether measured in the laboratory (peak oxygen uptake) or in the field (6-minute walk test); and 7) the degree of functional dyspnea. These characteristics can practically be integrated into multidimensional tools such as the BODE index capable of providing a more comprehensive evaluation of COPD subjects (34). The determination of these phenotypic characteristics is not only scientifically interesting, but also clinically important because they confer prognostic value and more importantly, they determine response to therapy. Although COPD genetic studies have focused primarily on the presence/absence of COPD, analysis of these additional phenotypes could provide useful insights into COPD pathophysiology.

The study of COPD phenotypes is relevant to disease etiology, pathophysiology, and treatment. The identification of clinically relevant phenotypes would change the present view of COPD as a unique multicomponent disease (35, 36) to a syndrome with multiple phenotypic expressions, thus changing (and challenging) current taxonomy of chronic airway diseases (37). Regarding disease etiology, the identification of non-genetic determinants of diseases will also benefit from an appropriate definition of phenotypes. It is also likely that the traditional approach to address heterogeneity (i.e., stratification by socio-demographic, clinical, or environmental factors) is likely to lead to a reduction in statistical power (30). On the other hand, the identification of clinically relevant phenotypes should also lead to increased understanding of the underlying pathobiology that contributes to a particular phenotype (31). Despite the huge advances in our understanding of the pathology of COPD in recent decades, there have been few attempts to link COPD pathologies to clinical COPD phenotypes (38). Finally, it has been hypothesized that failure to identify COPD phenotypes may limit the power of therapeutic trials (39) as effective and safe therapy is likely to differ across phenotypes (31, 40). Several
already existing examples illustrate this point, including the use of long-term oxygen therapy for COPD with chronic respiratory failure (but not for those with PaO2 values above 60 mmHg) (41, 42), the use of lung volume reduction surgery for patients with upper-lobe predominant emphysema and poor exercise capacity after rehabilitation (43), and, more recently, the development of roflumilast (a novel orally available anti-inflammatory drug) for only a subgroup of patients with COPD (those with chronic bronchitis) (44). Large, ongoing COPD studies may provide insight into phenotyping based on larger populations with more detailed descriptors.

Chest CT phenotypes
The use of chest CT scans for determination of lung density was first described in the 1980s as a measure of the degree of emphysema in COPD (45). An important step was the introduction of digital image analysis software, such that the density of the entire lung can now be reported as the lower 15th percentile in Hounsfield units or the percentage of lung density of the entire lung can now be reported as the lower 15th percentile in Hounsfield units or the percentage of lung density below a specific density mask threshold (e.g., < −950 HU) to define emphysema.

A more recent approach is the assessment of the thickness of the airway walls in order to determine the degree of airway remodelling. This was initially applied to asthma and more recently also to COPD (46). This approach appears to be robust for larger airways, and percent airway wall area can be used as a read-out (47). Although large airway dimensions correlate with small airway dimensions (48), direct assessment of the latter (airways <2 mm in diameter) is beyond the resolution of current CT scanning techniques. Quantitative assessment of chest CT scans for emphysema and airway disease provides an opportunity to define these two key phenotypes of COPD by objective criteria. There may also be further relevant CT-defined phenotypes that need more detailed study as to their clinical relevance and this includes emphysema distribution, emphysema pathological subtype (centrilobular vs. panlobular vs. paraseptal), the degree of mucous-mediated obstruction (plugging of airways), and bronchiec-
tasis.

An important issue for multicenter trials is standardization of CT measurements across different clinical centers. Here the different brands and models of CT scanners, which use different scanning technologies, scanning protocols, and different algorithms for data processing, can affect the lung density and the airway wall results. Careful standardization including the use of phantoms for all scanners is required in order to be able to compare results. In spite of these problems, it may be possible to obtain data on CT-assessed emphysema which could be compared in a multicenter study.

EXISTING COPD STUDY POPULATIONS
At the Boston meeting, 38 study populations which included spirometric assessment of COPD and DNA sample collection were reviewed (Table 1). These studies included 20 case-control studies (or studies of cases only or controls only, which will all be included as “case-control” for this discussion), 16 population-based cohort studies (some of which had family components), and two family-based studies. Despite the smaller number of studies, a much larger number of total subjects (>130,000) were available in the population-based cohort studies than in the case-control studies (approximately 38,000). The majority of studies have been performed in White populations. Most of the case-control studies include post-bronchodilator spirometry and a minimum number of pack-years of smoking criterion for inclusion, while most of the population-based cohort studies do not (Table 2). A surprisingly large fraction of case-control studies as well as some of the population-based studies included chest CT scan assessment. COPD exacerbations were also assessed in many studies.

Using the reported definitions of COPD and non-COPD from each study, there are approximately 39,600 COPD cases and 131,600 control subjects in the combined set of case-control and population cohort studies (Table 3). In these case-control and population cohort studies, there are approximately 14,700 cases and 37,600 controls reported to have genome-wide SNP genotyping currently available.

WORKSHOP RECOMMENDATIONS FOR CREATION OF AN INTERNATIONAL COPD GENETICS CONSORTIUM
Rationale and vision
At the Boston meeting, the participants identified multiple advantages of creating a COPD genetics research consortium, and strongly endorsed this approach. Larger sample sizes of cases and controls will definitely increase power to detect COPD susceptibility loci. The potential to assemble large numbers of severe COPD subjects that are clearly affected is a major advantage, since relatively small numbers of severely affected subjects have been included in most individual studies. Similarly, the opportunity to perform pooled analyses of chest CT phenotypes as seen in a major strength, as long as the technical challenges of different CT scanning and analytical protocols can be overcome. Opportunities to study other COPD-related phenotypes, including COPD exacerbation frequency, lung function decline, and lung cancer, were also recognized. Although many of the participating studies do not yet have genome-wide SNP genotyping, these studies provide opportunities to replicate initial GWAS findings in large numbers of additional subjects. In addition to studies of main genetic effects, a large COPD genetics consortium would improve the statistical power to study gene-by-environment interactions.

Although the studies listed above could be performed in a fairly short time-frame, potential future advantages of a COPD genetics consortium were also appreciated. Such a consortium could provide a framework for future genetic collaborations in exome sequencing and whole genome sequencing, as well as in other genetic/genomic areas (e.g., epigenetics, gene expression). There would likely be increased standardization of study protocols and procedures for future studies (e.g., imaging, questionnaires) and the potential for collaborative studies of non-genetic issues (e.g., phenotypes, biomarkers). Limiting duplication of research effort
<table>
<thead>
<tr>
<th>Study name</th>
<th>Study type</th>
<th>Current sample size*</th>
<th>Estimated final sample size for genetic studies</th>
<th>Genome-wide SNP genotyping platform</th>
<th>Country</th>
<th>Race/ethnicity</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANOLD**</td>
<td>Case-Only</td>
<td>1011</td>
<td>1,200</td>
<td>None</td>
<td>Nine Asian countries</td>
<td>All Asian</td>
<td>(56)</td>
</tr>
<tr>
<td>COPACETIC</td>
<td>Case-Control</td>
<td>1029</td>
<td>15,000</td>
<td>Illumina 610</td>
<td>Europe</td>
<td>All White</td>
<td>(57)</td>
</tr>
<tr>
<td>COPDGENE**</td>
<td>Case-Control</td>
<td>3590</td>
<td>10,000</td>
<td>Illumina omni-express</td>
<td>USA</td>
<td>2/3 non-Hispanic Whites (NHW) and 1/3 African Americans (AA)</td>
<td></td>
</tr>
<tr>
<td>COSYCONET**</td>
<td>Case-Control</td>
<td>3000</td>
<td>&gt;6000</td>
<td>Not determined</td>
<td>Germany</td>
<td>All White, America, Australia</td>
<td>(58)</td>
</tr>
<tr>
<td>ECLIPSE</td>
<td>Case-Control</td>
<td>1839</td>
<td>2043</td>
<td>Illumina 550</td>
<td>Europe</td>
<td>All White</td>
<td>(59)</td>
</tr>
<tr>
<td>EU COPD/East midlands consortium</td>
<td>Case-Control</td>
<td>1616</td>
<td>2528</td>
<td>Not determined</td>
<td>Europe</td>
<td>All White</td>
<td>(59)</td>
</tr>
<tr>
<td>EvA**</td>
<td>Case-Control</td>
<td>290</td>
<td>1,200</td>
<td>Illumina</td>
<td>Europe</td>
<td>All White</td>
<td>(60)</td>
</tr>
<tr>
<td>GenKOLS</td>
<td>Case-Control</td>
<td>933</td>
<td>1853</td>
<td>Illumina 550</td>
<td>Norway</td>
<td>All White</td>
<td>(61)</td>
</tr>
<tr>
<td>Hokkaido</td>
<td>Case-Control</td>
<td>274</td>
<td>500</td>
<td>None</td>
<td>Japan</td>
<td>All Asian</td>
<td>(62)</td>
</tr>
<tr>
<td>ICE COLD</td>
<td>Case-Only</td>
<td>411</td>
<td>411</td>
<td>Not determined</td>
<td>Switzerland/Netherlands</td>
<td>All White</td>
<td>(63)</td>
</tr>
<tr>
<td>KOLD**</td>
<td>Case-Control</td>
<td>321</td>
<td>700 COPD Cases</td>
<td>None</td>
<td>Korea</td>
<td>All Asian</td>
<td>(64)</td>
</tr>
<tr>
<td>Leuven</td>
<td>Case-Control</td>
<td>800</td>
<td>1,250</td>
<td>Not applicable</td>
<td>Belgium</td>
<td>All White</td>
<td>(65)</td>
</tr>
<tr>
<td>Lung Health Study</td>
<td>Case-Only</td>
<td>5887</td>
<td>4191</td>
<td>Illumina 660W</td>
<td>USA and Canada</td>
<td>4126 Whites, 65% AA</td>
<td>(66)</td>
</tr>
<tr>
<td>New Zealand</td>
<td>Case-Control</td>
<td>458</td>
<td>946</td>
<td>Not applicable</td>
<td>New Zealand</td>
<td>All White</td>
<td>(67)</td>
</tr>
<tr>
<td>NETT</td>
<td>Case-Only</td>
<td>366</td>
<td>366</td>
<td>Illumina 610</td>
<td>USA</td>
<td>All White</td>
<td>(68)</td>
</tr>
<tr>
<td>PAC COPD</td>
<td>Case-Only</td>
<td>342</td>
<td>342</td>
<td>0</td>
<td>Spain</td>
<td>All White</td>
<td>(69)</td>
</tr>
<tr>
<td>SPIROMICS**</td>
<td>Case-Control</td>
<td>0</td>
<td>3200</td>
<td>Not Determined</td>
<td>USA</td>
<td>Not Determined</td>
<td>(70)</td>
</tr>
<tr>
<td>Transcontinental:</td>
<td>Case-Control</td>
<td>80</td>
<td>600</td>
<td>None</td>
<td>Korea</td>
<td>All Asian</td>
<td>(71)</td>
</tr>
<tr>
<td>Transcontinental:</td>
<td>Case-Control</td>
<td>302</td>
<td>635</td>
<td>None</td>
<td>Poland</td>
<td>All White</td>
<td>(72)</td>
</tr>
<tr>
<td>Wake Forest</td>
<td>Case-Control</td>
<td>490</td>
<td>1,000</td>
<td>Illumina</td>
<td>USA</td>
<td>91% White 8% AA</td>
<td>(73)</td>
</tr>
<tr>
<td>Atherosclerosis Risk in</td>
<td>Pop-Based</td>
<td>1194</td>
<td>7545</td>
<td>Affymetrix 6.0 + IBC chip (Both Whites and African Americans)</td>
<td>USA</td>
<td>White</td>
<td>(74)</td>
</tr>
<tr>
<td>Communities (ARIC)*</td>
<td></td>
<td>247</td>
<td>1792</td>
<td></td>
<td>African American</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study/Population</td>
<td>Sample Size</td>
<td>Ethnicity</td>
<td>Platform</td>
<td>Country</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------</td>
<td>-------------</td>
<td>----------------------------</td>
<td>-----------------------------------------------</td>
<td>-----------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cardiovascular Health Study (CHS)</td>
<td>Pop-Based</td>
<td>USA White</td>
<td>370CNV beadchip + IBC chip (Both Whites and African Americans)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>402 292 2183 402 292 2183 2585</td>
<td>African American White</td>
<td>Illumina 370CNV beadchip + IBC chip (Both Whites and African Americans)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cleveland Family Study</td>
<td>Family/Pop-Based</td>
<td>USA White</td>
<td>Illumina 370CNV beadchip + IBC chip (Both Whites and African Americans)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>17 4 203 0 0 0 4 188</td>
<td>African American White</td>
<td>Illumina 370CNV beadchip + IBC chip (Both Whites and African Americans)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Copenhagen City Heart Population</td>
<td>Pop-Based</td>
<td>Denmark n/a</td>
<td>Affymetrix 6.0</td>
<td>Denmark n/a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Copenhagen General Population</td>
<td>Pop-Based</td>
<td>Denmark n/a</td>
<td>Affymetrix 6.0</td>
<td>Denmark n/a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coronary Artery Risk Development in Young Adults (CARDIA)²</td>
<td>Pop-Based</td>
<td>USA White</td>
<td>IBC chip</td>
<td>USA White</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>31 2 1148 0 0 0 1179</td>
<td>African American White</td>
<td>Illumina 370CNV beadchip + IBC chip (Both Whites and African Americans)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMCAP</td>
<td>Pop-Based</td>
<td>USA All White</td>
<td>Affymetrix 500 + 50K gene-centric chip</td>
<td>USA All White</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Framingham Heart Study</td>
<td>Family/Pop-Based</td>
<td>USA All White</td>
<td>Affymetrix 500 + 50K gene-centric chip</td>
<td>USA All White</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jackson Heart Study</td>
<td>Pop-Based</td>
<td>USA All White</td>
<td>Affymetrix 6.0 + IBC chip</td>
<td>USA All White</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lifelines</td>
<td>Family/Pop-Based</td>
<td>USA All White</td>
<td>Affymetrix 6.0 + IBC chip</td>
<td>USA All White</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lovelace Smokers Cohort</td>
<td>Pop-Based</td>
<td>USA All White</td>
<td>Affymetrix 6.0 + IBC chip</td>
<td>USA All White</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multi-Ethnic Study of Atherosclerosis (MESA) Lung Study²</td>
<td>Pop-Based</td>
<td>USA All White</td>
<td>Affymetrix 6.0 + IBC chip</td>
<td>USA All White</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MESA Lung SHARE</td>
<td>Family/Pop-Based</td>
<td>USA All White</td>
<td>Affymetrix 6.0 + IBC chip</td>
<td>USA All White</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Continued on next page)
Table 1. International COPD Genetics Consortium study populations—Genetics (Continued)

<table>
<thead>
<tr>
<th>Study name</th>
<th>Study type</th>
<th>Current sample size*</th>
<th>Genome-wide SNP genotyping availability</th>
<th>Estimated final sample size for genetic studies</th>
<th>Genome-Wide SNP genotyping platform</th>
<th>Country</th>
<th>Race/ethnicity</th>
<th>Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normative Aging Study</td>
<td>Pop-Based</td>
<td>191</td>
<td>0</td>
<td>1215</td>
<td>Illumina 610</td>
<td>USA</td>
<td>All White</td>
<td>(17,27)</td>
</tr>
<tr>
<td>Rotterdam Study (RS1, RS2 and RS3 combined)*</td>
<td>Pop-Based</td>
<td>209</td>
<td>209</td>
<td>2881</td>
<td>Illumina Infinium II 550K</td>
<td>Netherlands</td>
<td>All White</td>
<td>(15)</td>
</tr>
<tr>
<td>SpiroMeta**</td>
<td>Pop-Based</td>
<td>3278</td>
<td>356</td>
<td>63,079</td>
<td>Multiple GWAS platforms (n = 24, 756); multiple non-GWAS platforms (n = 38, 314)</td>
<td>Europe and Australia</td>
<td>All White</td>
<td>(16)</td>
</tr>
<tr>
<td>Vlagtwedde/Vlaardingen</td>
<td>Pop-Based</td>
<td>351</td>
<td>27</td>
<td>1342</td>
<td>None</td>
<td>Netherlands</td>
<td>All White</td>
<td>(70)</td>
</tr>
<tr>
<td>Boston Early-Onset COPD</td>
<td>Family</td>
<td>200 Probands</td>
<td>0</td>
<td>1200</td>
<td>None</td>
<td>USA</td>
<td>98% White</td>
<td>4</td>
</tr>
<tr>
<td>International COPD Genetics Network</td>
<td>Family</td>
<td>983 Probands</td>
<td>0</td>
<td>2859</td>
<td>None</td>
<td>North America and Europe</td>
<td>Mostly White</td>
<td>(71)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>~1000 Relatives</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note that “All COPD” and “Control” were defined differently within each study and/or consortium. “Severe COPD” was defined as GOLD 3–4 using post-bronchodilator spirometry if available, and pre-bronchodilator spirometry if not available. For the CHARGE consortium studies, severe COPD was defined as noted below:

**Indicates that study population is still being recruited.

*Participating in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium which includes only Whites from these cohorts. In CHARGE, COPD was defined as FEV1 < LLN and FEV1/FVC < 0.7. Controls were defined by FEV1 and FVC all > LLN.

#Participating in the Candidate Gene Association Resource (CARe) consortium (72). For African Americans in CARe, “COPD” was defined as pre-BD ratio < 0.70 and FEV1%pred < 80%; “severe COPD” as pre-BD ratio < 0.70 and FEV1%pred < 50%; and controls as pre-BD ratio > 0.70 and FEV1%pred > 80%. Note that for cohorts participating in both CHARGE and CARe, definitions of COPD, severe COPD, and control status used in this table differ for Caucasians and African-Americans from the same participating cohort. Additionally, for the ACR cohort, case and control counts for the total (“current”) sample of African-Americans are based on the CHARGE definitions rather than the CARe definitions.

‡SpiroMeta—Cases: FEV1 < 80% predicted and FEV1/FVC < 0.7; Controls: FEV1 > 80% predicted and FEV1/FVC ratio of > 0.7. Numbers of individuals known to be cases and known to be controls are shown (further data collection is underway; the number of known COPD cases and documented controls will increase when this information is received).

Characterized with CT scan only.
## Table 2. International COPD Genetics Consortium study populations—Phenotypes

<table>
<thead>
<tr>
<th>Study</th>
<th>Age Range</th>
<th>Gender (% Male)</th>
<th>Minimum Pack-Years</th>
<th>Pre-BD</th>
<th>Post-BD</th>
<th>Obtained</th>
<th>Visual Assessment</th>
<th>Quantitative Assessment</th>
<th>Exac**</th>
<th>Co-Mor**</th>
<th>Lung Function Decline</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANOLD</td>
<td>40–91</td>
<td>94</td>
<td>10</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
<td>–</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>COPACETIC</td>
<td>50–80</td>
<td>98</td>
<td>20</td>
<td>Yes</td>
<td>Some</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>COPDGene</td>
<td>45–80</td>
<td>52</td>
<td>10</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>COSYCONET</td>
<td>&gt;40</td>
<td>n/a</td>
<td>0</td>
<td>Yes</td>
<td>Yes</td>
<td>Some</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>ECLIPSE</td>
<td>40–75</td>
<td>66</td>
<td>10</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>EU COPD/East Midlands Consortium</td>
<td>Mean 66</td>
<td>70</td>
<td>10</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
<td>–</td>
<td>No</td>
<td>Yes</td>
<td>Some</td>
</tr>
<tr>
<td>EvA</td>
<td>45–75</td>
<td>65</td>
<td>5</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>GenKOLS</td>
<td>40–81</td>
<td>66</td>
<td>2.5</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Hokkaido</td>
<td>41–87</td>
<td>94</td>
<td>12</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>ICE COLD</td>
<td>Mean 66</td>
<td>57</td>
<td>0</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
<td>–</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>KOLD</td>
<td>45–82</td>
<td>92</td>
<td>10</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Leuven</td>
<td>43–90</td>
<td>75</td>
<td>15</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Lung Health Study</td>
<td>35–60</td>
<td>63</td>
<td>10</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
<td>–</td>
<td>Yes</td>
<td>Some</td>
<td>Yes</td>
</tr>
<tr>
<td>New Zealand</td>
<td>40–85</td>
<td>60</td>
<td>15</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>–</td>
<td>–</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>NETT</td>
<td>Mean 67</td>
<td>73</td>
<td>0</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>PAC COPD</td>
<td>Mean 68; Range 44–86</td>
<td>93</td>
<td>0</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>SPIROMICS</td>
<td>TBD</td>
<td>TBD</td>
<td>TBD</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Transcontinental: Korea</td>
<td>40–80</td>
<td>97</td>
<td>10</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
<td>–</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Transcontinental: Poland</td>
<td>40–80</td>
<td>69</td>
<td>10</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
<td>–</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Wake Forest</td>
<td>42–88</td>
<td>96</td>
<td>20</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>–</td>
<td>–</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>ARIC</td>
<td>45–64</td>
<td>45</td>
<td>0</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>–</td>
<td>–</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Cardiovascular Health Study</td>
<td>65–95</td>
<td>43</td>
<td>0</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>–</td>
<td>–</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Cleveland Family Study</td>
<td>23–88</td>
<td>48</td>
<td>0</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>–</td>
<td>–</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Copenhagen City Heart</td>
<td>n/a</td>
<td>n/a</td>
<td>0</td>
<td>Yes</td>
<td>n/a</td>
<td>No</td>
<td>–</td>
<td>–</td>
<td>Yes</td>
<td>Yes</td>
<td>n/a</td>
</tr>
<tr>
<td>Copenhagen General Population</td>
<td>n/a</td>
<td>n/a</td>
<td>0</td>
<td>Yes</td>
<td>n/a</td>
<td>No</td>
<td>–</td>
<td>–</td>
<td>Yes</td>
<td>Yes</td>
<td>?</td>
</tr>
<tr>
<td>CARDIA</td>
<td>27–44</td>
<td>45</td>
<td>0</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>–</td>
<td>–</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>EMCAP</td>
<td>&gt;60</td>
<td>50</td>
<td>10</td>
<td>Yes</td>
<td>Some</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Framingham</td>
<td>19–92</td>
<td>46</td>
<td>0</td>
<td>Yes</td>
<td>Some</td>
<td>In Subset</td>
<td>Pending</td>
<td>Pending</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Jackson Heart Study</td>
<td>21–93</td>
<td>38</td>
<td>0</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>–</td>
<td>–</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Lifelines</td>
<td>21–88</td>
<td>41</td>
<td>0</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>–</td>
<td>–</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Lovelace Smokers Cohort</td>
<td>40–75</td>
<td>22</td>
<td>15</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
<td>–</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>MESA Lung</td>
<td>45–84</td>
<td>49</td>
<td>0</td>
<td>Yes</td>
<td>Some</td>
<td>Yes</td>
<td>Pending</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Pending</td>
</tr>
<tr>
<td>MESA Lung SHARE</td>
<td>45–84</td>
<td>50</td>
<td>0</td>
<td>No</td>
<td>No</td>
<td>Yes (cardiac)</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Normative Aging Study</td>
<td>21–&gt;80</td>
<td>100</td>
<td>0</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>–</td>
<td>–</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Rotterdam Study</td>
<td>&gt;45</td>
<td>45</td>
<td>0</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>–</td>
<td>–</td>
<td>No</td>
<td>Yes</td>
<td>Pending (RS1)</td>
</tr>
<tr>
<td>SpiroMeta</td>
<td>8–93*</td>
<td>42*</td>
<td>0</td>
<td>Yes</td>
<td>In a Subset</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Vlagtwedde/ Vlaardingen</td>
<td>35–79</td>
<td>51</td>
<td>0</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>–</td>
<td>–</td>
<td>No</td>
<td>Some</td>
<td>Yes</td>
</tr>
<tr>
<td>Boston Early-Onset COPD</td>
<td>&gt;10</td>
<td>42</td>
<td>0</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>–</td>
<td>–</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>International COPD Genetics Network</td>
<td>45–65</td>
<td>46</td>
<td>5</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

*For a subset of individuals (n = 20, 288) with information on age and gender currently available.
**Abbreviations: "Exac" = Data available regarding COPD exacerbations; "CoMor" = Data available regarding co-morbid diseases.
could likely be accomplished as well. Overall, there was general consensus that a COPD genetics consortium would have a high likelihood of significantly advancing knowledge in the field.

In addition to these advantages of forming a collaborative consortium, a variety of challenges were identified. It was recognized that there are academic realities including the need for individual research groups to demonstrate academic productivity to renew funding and promote research personnel. Some COPD genetics collaborations already exist, and a goal was not to interfere with those existing relationships. Although studies that include reasonable numbers of COPD cases and control subjects could be analyzed individually and combined using meta-analytical approaches, the optimal approach for utilizing studies of COPD cases only or controls only was not as clear.

A variety of challenges related to phenotypic characterization were also identified. Substantial variation exists in the definitions of cases and controls between studies (e.g., physiologic measurements of lung function using GOLD criteria or use of lower limit of normal [LLN]), as well as in spirometry protocol (e.g., pre- vs. post-bronchodilator). Some phenotypes (e.g., imaging) may be difficult to combine across studies due to technical issues. There are important variations in study populations (e.g., race/ethnicity, smoking history, exclusion of subjects with other illnesses, other criteria used for selection, study design, and informed consent restrictions) and genetic analysis approaches (e.g., variation in genotyping platform, data cleaning, analytical approaches, and data sharing).

Despite these challenges, there was general agreement that the advantages of collaboration far outweighed the limitations, and that a transparent and open collaboration could overcome most of the challenges. To be successful, the needs and rights of each contributing study will need to be respected. Based on the enthusiastic support for an international COPD genetics consortium from the Boston meeting participants, the research projects amenable to this consortium approach and an organizational structure for the consortium were discussed.

### Table 3: International COPD Genetics Consortium: Approximate current study population sample sizes

<table>
<thead>
<tr>
<th></th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case-Control Studies</td>
<td>23,039</td>
<td>15,610</td>
</tr>
<tr>
<td>Cohort Studies</td>
<td>16,526</td>
<td>115,956</td>
</tr>
<tr>
<td>Combined Set</td>
<td>39,565</td>
<td>131,566</td>
</tr>
<tr>
<td>GWAS in Combined Set</td>
<td>14,741</td>
<td>37,612</td>
</tr>
</tbody>
</table>

Notes: For this table, approximate sample sizes from Table 1 were used as actual values. Only subjects with spirometry from Table 1 were included in these calculations (MESA Lung SHARe was excluded). Population-based studies with a family component (e.g., Framingham) are listed under “Cohort Studies,” while other pedigree-based studies (e.g., International COPD Genetics Network) have been excluded. For the “GWAS in Combined Set Currently Available” only subjects with available genotyping by December 2010 were included.

### Feasibility of collaborative COPD genetics studies

Although the development of large consortia of thousands of subjects may obviate some of the issues that have contributed to non-replication of previous COPD genetic studies (such as power limitations germane to smaller studies), the inclusion of data from a large number of studies presents unique challenges and opportunities.

### Smoking exposure and penetrance

Despite the challenges of disease gene discovery in complex disease, there are some striking advantages to studying the genetics of COPD (COPD strictly being a syndrome not a specific disease). First, one of the most important features of studying the genetics of COPD is that the key environmental exposure of cigarette smoking is known and quantifiable in the setting of a gene-by-environment interaction. In contrast to most other complex diseases, the majority of COPD can be attributed to a single exposure (cigarette smoking) which can be crudely quantified, by intensity (cigarettes/day) and/or total exposure (pack-years), across both cases and controls in geographically diverse populations. The central role of smoking exposure in genetic susceptibility is illustrated by the divergent outcomes in people with alpha-1 antitrypsin deficiency based on their smoking history (49).

Second, although COPD is a syndrome encompassing both emphysema and small airway disease that are present in varying degrees, both are characterized by irreversible airflow limitation (reduced FEV1 and FEV1/FVC ratio), which can be measured by simple spirometry in population studies. From an epidemiological perspective, FEV1 (after age, gender, race, and height adjustment) provides a good starting place from which to define COPD, as it is a highly heritable trait (50) regardless of the heterogeneity of COPD subjects. Moreover, with increasing smoking exposure, FEV1 defines susceptible and resistant smokers with an increasingly bimodal distribution supportive of a genetic basis (2, 51–53) and possibly a threshold effect. Comparing smokers at either end of the FEV1 spectrum but with comparable smoking exposure, so called “extreme phenotypes” (54), may help to overcome minor differences in spirometric criteria defining the COPD phenotype.

Despite these characteristics of COPD as a complex genetic disease, there remain significant challenges in combining population-based and case-control samples. Many studies that can be included in a collaborative COPD meta-analysis have not taken detailed smoking histories or validated intensity of current smoking via measurements of cotinine levels. Although reporting bias is a concern, self-report of cigarette smoking has been demonstrated as a reliable assessment.

Some of the studies proposed for inclusion in this consortium have focused on a minimum amount of smoking exposure for enrollment, while others have not. Similarly, some studies have focused on the heavy smoker, and some have included a range of exposures. Including all studies allows for a reasonable attempt to achieve the necessary power to assess genetic main effects, but also gene-by-smoking interactions through stratification and/or adjustment. In addition, genetic insights into COPD will be gleaned by not only studying those genes that associate with COPD susceptibility, but also genes that may portend protective resistance to COPD.
in subjects with an extremely high number of pack-years but normal lung function.

**Heterogeneity of COPD**

First and foremost in the planning and organization of large consortia with the goal of meta-analysis, the heterogeneity of phenotypes across studies needs to be addressed. This is an issue by no means limited to respiratory genetic studies. A paramount challenge in studies of COPD has been the inherent heterogeneity of the disease, variable effects of smoking exposure on penetrance (described previously), and the importance of defining disease subtypes. In addition, not all studies have performed pre- and post-bronchodilator spirometry, and many studies have not been sufficiently resourced to undertake chest CT scanning to phenotype COPD pathologically. Even in the presence of post-bronchodilator spirometry, issues of spirometric diagnosis of COPD based on using GOLD criteria versus lower limit of normal may contribute to phenotypic heterogeneity. This is likely to be minimized in those case-control studies comparing more extreme (susceptible vs resistant) phenotypes where misclassification of cases or controls based on variation in spirometry-based definition is likely to be minor. However, severity of COPD is an issue that needs careful consideration given the strong association between aging and loss of lung function. Moreover, it is noted that a spirometric diagnosis of a resistant smoker does not obviate some misclassification as smoking-related emphysema may be present despite normal spirometric measures.

A major strength of this consortium will be the availability of data from both spirometry and CT scans of the lungs for parsing subjects by emphysema status (and also within COPD cases for emphysema versus airway predominant disease). Although CT scanning presents challenges (lack of uniformity of scanner technical characteristics, scanning protocols, radiation dosing, scoring of emphysema severity/distribution, etc.), computerized approaches to process and analyze chest CT scans may be useful in harmonizing CT scan data, and will likely assist in refinement of COPD subtypes.

**Ethnic heterogeneity/population substructure**

Although the inclusion of data from Caucasian, African, Native American, and Asian subjects may lead to false positive and/or false negative association findings due to population substructure, contribution of all subjects’ data to the power of the overall analysis, and to the race-specific genetic association analyses of COPD, are major strengths of performing a multi-ethnic consortium. There are many existing examples of disease associations confined to specific ethnic groups. Primary analyses would be conducted within each ethnic group, followed by comparison of association results between groups.

**Study-specific issues**

Although the inclusion of many studies of COPD may increase the power due to an increase in total subject number, presently a minority of COPD studies have GWA data. Given the large burden of disease accounted for by COPD, this void of GWA data in and of itself supports the importance of efforts to accrue more GWA data. The consortium will include GWA data on both population-based cohorts and case-control studies. In the population-based cohorts, the contribution of genetic variants on lung function can be explored in settings other than chronic smoking.

There is a need for more GWA data in case-control studies, where smoking has been accounted for and other important exposures have been examined. Case-control studies can also allow investigation of the role in disease of variants shown to be associated with lung function in population-based studies. The case-only studies can be added to the case-control studies in a mega-analysis of individual genotype and phenotype data, or in circumstances where data can be combined or compared with the controls derived from other studies. The consortium is also fortunate enough to have large prospective studies in which the genetic determinants of rate of decline in lung function (or other aspects of disease progression) can be studied; this could be insightful, for example, for those variants shown to be associated with COPD or cross-sectional lung function measures.

Other study-specific issues include variations in enrollment criteria, age ranges of subjects, and rates of co-morbid conditions including obesity which may affect lung function. Of these, variations in smoking history (described above) and current smoking status are potentially the most essential, as some of the COPD studies had minimum amounts of smoking exposure required for eligibility, and include a mix of current and former smokers. Given that gene-by-smoking interactions are crucial to include in genetic analyses (2, 51–53) this variable inclusion of subjects may seem to be a limitation.

For genetic studies there is likely enrichment in genetic effects in those subjects who develop COPD at a very young age as well as those smokers who remain resistant to both COPD and emphysema at very old ages. Associations will need to be re-examined with stratification by age of disease onset, total pack-year exposure and current smoking status where the data are known.

Variable rates of comorbidities in the different COPD studies may impact genetic associations with lung function (such as the association of diabetes with lower lung function) but the inclusion in genetic analysis of the most diverse group with COPD may increase the likelihood that positive associations are true positive findings.

**GWAS platforms and data cleaning**

As has been the challenge in other complex diseases such as diabetes, the platforms used for genome-wide genotyping have varied. There is variable inclusion of SNPs leading to differential coverage of genes on a given platform. However, this reality has led to the availability of imputation methods to overcome the differences between GWA arrays; these novel *in silico* tools allow for the development of a larger study population less constrained by choice of genotyping technology. In addition to the genotyping platform, approaches to
Sharing of individual subject data
The protection of human data and subject privacy is paramount and the ability to share individual level genotype results may be limited. Thus, the performance of "mega-analysis" in which individual-level genotype and phenotype data would be shared, though remaining a worthwhile eventual goal, was judged not to be essential to progress at this time. However, meta-analytic approaches that use study-wide association data (p-values) weighted by study size or by inverse variance have been shown to be as powerful as mega-analysis approaches that utilize subject level data (55). Thus for identification of common variants for COPD (at least 5% minor allele frequency) meta-analytic approaches will provide important insights into COPD.

Plan for initial meta-analysis
A preliminary design for the initial collaborative genetic association meta-analysis for the consortium was created at the Boston meeting. Two key genome-wide association analyses were proposed: 1) All COPD vs. Controls, and 2) Severe COPD vs. Controls. The precise definitions of All COPD and Severe COPD remain to be determined. Within each study population that has existing genome-wide SNP genotyping data, standard quality control approaches will be used to clean the data, including criteria for exclusion of SNPs with low call rates, low minor allele frequency, departures from Hardy-Weinberg equilibrium, and differential rates of missing data between cases and controls, and exclusion of individuals with low call rates or exhibiting cryptic relatedness among unrelated samples.

Standard approaches to genotype imputation will be applied in each study, followed by a similar approach to population stratification adjustment within each study using adjustment with principal components for genetic ancestry. Genome-wide association analysis for the two COPD affection status phenotypes (all COPD and severe COPD) will be performed within each study, with separate analyses in subjects of Caucasian, Asian, and African ancestry. Meta-analysis of GWAS will be performed within each major racial group using inverse variance weighted meta-analysis methods to account for differences in sample size and imputation quality across genotyping platforms, followed by comparison of association evidence between major racial groups. Finally, replication genotyping and association analysis of the most interesting SNPs will be performed in the remaining study populations without genome-wide SNP data.

Structure of the Consortium
The mandate of the International COPD Genetics Consortium is to find common and rare genetic determinants of COPD; to identify COPD subtypes and their genetic basis; and to use this information to develop new disease classifications and therapeutic interventions. Based on the discussions at the Boston meeting, it was recommended that research studies including COPD and control subjects would be invited to participate if they collected high quality spirometry data and DNA samples, and if the study met a minimum sample size. The expectation is that the studies will include at least 200 COPD cases and 200 controls, but review of specific studies is possible if those criteria are not met. For studies that include case-only collections, they would be encouraged to find appropriate sets of control subjects for genetic association analysis; if not available, those COPD study populations could be included in studies of COPD progression or CT subtypes. Study populations meeting these criteria that were not represented at the Boston meeting will be welcome to join this international collaborative effort.

Several committees will be created to perform the consortium research and administration, including a Steering Committee (in charge of major decisions); Planning/Executive Committee (routine operations); Phenotype Harmonization; Imaging Committee; Genotyping and Genomics Core; and Analysis Core.

The International COPD Genetics Consortium has the potential to provide short-term results by providing highly powered genome-wide association studies of COPD susceptibility, and long-term results by facilitating the study of other COPD-related phenotypes and other genomic outcomes. Organization, resources, and communication will be essential to realize this potential.

SUMMARY RECOMMENDATIONS
(1) Create the International COPD Genetics Consortium (ICGC) – to be open worldwide to include all study populations meeting minimum criteria for size, spirometric data, and DNA availability.
(2) Mandate of the ICGC is to:
(a) Use pooled resources to define rare and common genetic determinants of COPD
(b) Identify COPD subtypes and their genetic basis
(c) Develop new disease classifications for COPD
(d) Foster development of new therapeutic interventions that are subtype or disease classification specific
(3) Recommended committee structure:
(a) Steering Committee (with oversight of major decisions)
(b) Planning/Executive Committee
(c) Imaging Committee
(d) Phenotype Harmonization Committee
(e) Genotyping and Genomics Core
(f) Analysis Core
(4) Generating new GWAS/genotyping/sequencing/gene expression data
(a) Expand and extend existing and ongoing genetic analysis projects
(5) Plans for genetic analysis
(a) Initial meta-analysis focused on common definitions of case status and on extreme phenotypes
(b) Common standardized quality control approaches to clean data
(c) Standard approach for data analysis
REFERENCE


COPD: Journal of Chronic Obstructive Pulmonary Disease


