COPDGene® Central Data, Imaging and Human Subject Contacts Protocol

Version 6
October 6, 2014

Table of Contents:

Overview of the COPDGene Study

COPDGene® is a multi-institutional, longitudinal, NHLBI funded study of the genetic epidemiology of COPD. Following informed consent at the local clinical centers, subjects were enrolled in Phase 1 (2007-2012) of the study. In Phase 2 (2013 -2017) of the study, subjects are invited to return for a second visit and a new informed consent process is initiated for this second phase of the study. After the new informed consent is signed and the study visit completed, the appropriate specimens, CT Images, and clinical data are obtained for delivery to the Biorepository, Pulmonary Function Core, COPDGene® Quantitative Imaging Laboratory (QIL) and Data Coordinating Center (DCC) at National Jewish Health, respectively. During the first five years COPDGene® enrolled and characterized 10,364 subjects. Subjects will continue reporting exacerbations and other health information twice a year. The study will also enroll an additional group of non-smoker control subjects of up to 1500 individuals to define the impact of normal aging on physical function and CT chest imaging.

COPD is the third leading cause of death in the United States and as such, is a major public health issue. COPD is a heterogeneous condition with subtypes that have different rates of progression and clinical manifestations. These subtypes are likely to have different genetic factors and imaging characteristics. Phase 2 COPDGene emphasizes tracking the longitudinal change in subject's health to correlate with imaging characteristics and genetics.

A key characteristic of COPD is the progression to more severe disease and ultimately death. This progression is associated with increased needs for assistance and reduced capacity for independence. A direct consequence of functional declines is that sicker subjects are likely to move to assisted living situations or nursing home placement. Thus the most severely affected subjects are more likely to change address and other contact information, leading to inadvertent loss of contact from the study. Death is an important endpoint for COPDGene and is another common cause for loss of contact with study subjects. Effective longitudinal subject tracking and death tracking are critical to understand the natural history and subtypes of disease.

Because COPDGene is a genetic epidemiology study, Phase 1 of the study generated genome wide data for each of the study subjects. In phase 2 more detailed genetic sequencing of each subject is planned.

Background and Significance

Chronic obstructive pulmonary disease (COPD) is the third leading cause of mortality in the United States (1). Genetic studies of complex diseases like COPD have the potential to provide insight into the pathophysiological mechanisms of COPD susceptibility and heterogeneity. A strong genetic basis for the susceptibility of smokers to develop COPD is suggested by:

(1) Marked variability in the development of airflow obstruction among smokers (2);

(2) Clear familial clustering of COPD and COPD-related phenotypes (3); and

(3) Linkage of COPD-related phenotypes to specific genomic regions in families with severe, early-onset COPD (4).

Case-control studies have been performed for many candidate genes in COPD, but the results have been inconsistent (5). Possible contributors to these inconsistent results include:

1) small sample sizes; 2) inadequate classification of distinct phenotypes (e.g., emphysema vs. airway disease); 3) widely varying criteria used for case definition and control selection in different studies; 4) failure to assess (and, if necessary, adjust) for population stratification; 5) testing a limited number of genetic variants in each candidate gene; 6) genotyping error; and 7) lack of correction for multiple statistical testing. Recent progress in single nucleotide polymorphism (SNP) genotyping and DNA sequencing allows for association studies on a genome-wide scale, rather than limiting analysis to recognized candidate genes or regions of linkage; however, the multiple statistical tests involved in genome-wide association (GWA) studies of thousands of SNPs raise challenges in separating true from false positive associations (6). In addition, genetic association studies within a single racial/ethnic group may not generalize to other populations.

To address the multiple testing and generalizability problems of GWA studies, we have performed comprehensive GWA, exome sequencing and whole genome DNA sequencing to identify genes influencing COPD in two major racial/ethnic groups (non-Hispanic Whites and African Americans). In Phase 2 this work will be expanded to all subjects and additional information about longitudinal change will provide important new phenotypic markers for genetic analysis particular in discerning subjects with stable or slowly progressive disease compared to those with rapid progression.

Adding to the cross-sectional subject data collected in the baseline COPDGene study visit, disease progression and incidence of COPD in smokers are important additional endpoints for genetic association studies. We have completed enrollment of 10,300 subjects as the first phase of this project and now propose to invite these subjects to return for a second evaluation five years after the initial visit to assess disease incidence and/or progression in COPDGene subjects.

The primary goals of COPDGene Phase 2 are: 1) to identify new genetic loci that influence the development and/or progression of chronic obstructive pulmonary disease (COPD) and COPD-related phenotypes, and 2) to reclassify COPD into subtypes that can ultimately be used to develop effective subtype-specific therapies and prevention. The reclassification of COPD will be done using imaging, clinical and physiologic characteristics, longitudinal progression, long-term outcomes, and genetics.

In addition to identifying COPD genetic determinants, this program will characterize the natural history of COPD and identify well-characterized COPD subtypes. Improved understanding of COPD subtypes and genes controlling susceptibility to COPD could lead to novel patho-physiological insights, refined diagnostic criteria, and new treatment approaches. Moreover, the availability of comprehensive genetic data and longitudinal data on a large biracial group of smokers will be an invaluable national resource for other investigators.

Study Design for the Central COPDGene Study

COPDGene® Study has recruited 10,364 subjects stratified by severity of COPD, airway restriction, and smoking status (smoking controls - history of cigarette smoking with normal spirometry, non-smoking controls – no smoking history with normal spirometry) to conduct cross-sectional case-control studies. COPDGene has identified and phenotyped COPD cases and control subjects from two racial groups (non-Hispanic whites and non-Hispanic African Americans) for genetic, epidemiologic, and natural history studies. The COPDGene® cohort is comprised of current or former smokers, enrolled between the ages of 45 and 80 years, with a

minimum ten pack-year smoking history, as well as a small group of similar aged non-smokers. Pulmonary function testing is performed on all subjects. Subjects with COPD or evidence of decreased lung function are classified by GOLD stage. Subjects whose lung function is normal are assigned as controls. High-resolution chest CT scans were acquired during the study visit and later analyzed using Slicer and Vida Diagnostics PW2 analysis software. Blood was collected for DNA extraction and GWAS analysis. Questionnaires were administered and data has been collected. Data analysis from Phase 1 is ongoing. As specified in the NHLBI grant award a data sharing plan has been implemented and de-identified data has been made available through dbGaP http://www.ncbi.nlm.nih.gov/gap.

In addition to the cross-sectional initial data collection, subjects consented to be re-contacted either twice a year or up to four times a year to gather information about new onset of health conditions and the occurrence of respiratory exacerbations. In phase 1 they were actually contacted twice a year for follow-up surveys. These twice yearly reports are also stored in the Data Coordinating Center.

COPDGene® has included the collection of data from 21 distinct clinical centers: National Jewish Health, Ann Arbor VA Medical Center, Baylor College of Medicine, Brigham and Women's Hospital, Columbia Univ. Medical Center, Duke Univ. Medical Center, Fallon Clinic, Johns Hopkins University, L.A. Biomedical Research Inst, Michael E. DeBakey VAMC, Minneapolis VA Med Ctr, Morehouse School of Medicine, Univ. of Iowa, Univ. of Michigan, Temple University, Univ. of Minnesota, Univ. of Pittsburgh, UTHSC at San Antonio, Univ. of Alabama, Birmingham, Univ. of California, San Diego, Minnesota HealthPartners – Twin Cities.

This study protocol describes the central structure of the study, data handling procedures, research procedures to be carried out centrally and the procedures for the twenty one clinical centers and human subject research being performed under the auspices of the COPDGene main study.

Structure of the Central COPDGene Study

The primary purpose of the COPDGene® Central Data Management is to oversee and manage the study, provide a mechanism to maintain the longitudinal contact with study subjects, assure protection of participant privacy and ensure responsible handling of data. In Phase 2 of the study we will be collecting subject identifiers (name, address, phone numbers and social security numbers) and storing these in our Data Coordinating Center (DCC) located at National Jewish Health. In creating this Phase 2 DCC that will include subject identifiers we are seeking IRB review and approval of this human subject research. In Figure 1 we illustrate the basic structure of data flow within the study and note data transfers of protected health information from our subjects that include direct and indirect identifiers.

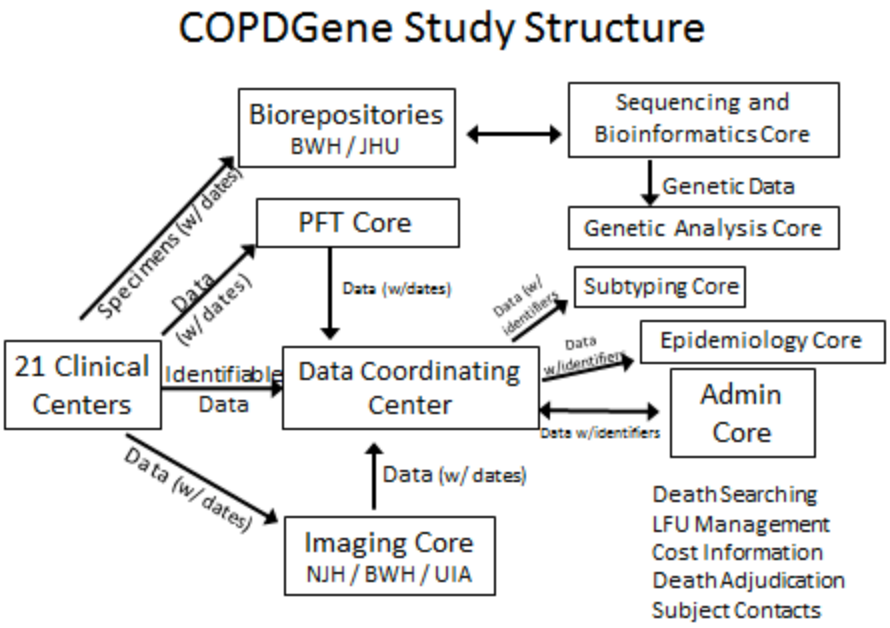

Figure 1. The structure of data handling and transfer for the COPDGene Study, Phase 2 between clinical centers and the study "Cores". Types of data (identifiable, with indirect identifiers and genetic data) are specified with the planned transfers.

The COPDGene study has a central Administrative Core that coordinates the study activities and works closely with the Data Coordinating Center to write and communicate the study protocol, consent template, manual of operations and to oversee training of the clinical center staff involved in the study. The Administrative Core also manages investigators meetings, provides scheduling and coordination for committee meetings and writing groups, oversees the longitudinal follow-up program, manages acquisition of Medicare cost information, performs Social Security Death Masterfile (SSDMF) searching and maintains the study files for death adjudication.

The Imaging Core receives subject CT scan image files and clinical center phantom data files along with dates of CT scan. The Imaging Core assesses the quality of the CT images and performs analysis of quantitative characteristics on the image files as well as overseeing the more complex scoring of images for visual analysis. Data and image files are shared between the three different sites (National Jewish Health, University of Iowa and Brigham and Women's Hospital) under the Imaging Core. Final output from the image analysis work is transferred and stored in the DCC

The Pulmonary Function Testing (PFT) Core receives subject data with a date of study visit from the twenty one clinical centers via a file depot located at National Jewish provided by the Data Coordinating Center. Dr Robert Jensen is the head of the PFT Core. He reviews the Spirometry and DLCO data for quality assessment and returns this data with a quality grade to the Data Coordinating Center. A back up data handling process for PFT quality assurance is located at National Jewish Health with Dr Rand Sutherland providing quality grading.

The Sample Storage Cores (Biorepositories) for COPDGene are located at Johns Hopkins University (JHU) and Brigham and Women's (BWH) Hospital. In phase 1 of the study biospecimens and DNA were stored at JHU, and in phase 2 of the study, biospecimens will be stored at BWH. Backup storage of DNA from phase 1 is already in place at BWH. The date of sample collection is part of the information stored at each site. All samples are otherwise stored with a code but no other identifiers. The coded identifier for both sites is the COPDGene subject ID number.

The Sequencing and Bioinformatics Core is located at Brigham and Women's Hospital and the Genetic Analysis Core is located at Brigham and Woman's Hospital, University of Colorado Denver and Johns Hopkins University. These Cores receive DNA samples for analysis and resultant genetic data for analysis. Samples and data are coded and have no other identifiers associated with them.

The Epidemiology Core is located at the University of Colorado Denver on the Anschutz campus in Aurora. The Epidemiology Core participates in data analysis of phenotypes, genetic data, and imaging data. These data will include subject identifiers - dates and geocodes.

The Subtyping Core is made up of investigators from the whole study who meet by teleconference and collaborate on data analysis. Data shared to this Core will include geocodes and location to the level of zip codes.

The Biomarkers Core is located at National Jewish Health under the oversight of Dr. Russell Bowler but works collaboratively with the Sample Storage (Biorepository) Core sharing biospecimens to be used for analysis of biomarkers. Data will include dates of sample collection.

The Mortality Adjudication Core is closely linked to the Administrative Core. They will receive death certificates, subject medical records and when the subject has consented to have personal identifiers transmitted to the Data Coordinating Center, staff for the Mortality Adjudication Core will contact next of kin for informant interviews and request death certificates.

The Principal Investigators for the study, the head of the Data Coordinating Center, the director of the Quantitative Imaging Lab, the Director of the Sequencing Center and Biorepository, Head of the Epidemiology Core, Head of Genetic Analysis Core, the Clinical Centers Director and the Associate Director for COPDGene, in consultation with the COPDGene Executive Committee, manage the central functions of the study. This team and the COPDGene® Investigators seek to encourage appropriate collaborative relationships with outside investigators to advance scientific knowledge and maximize the value of the study.

Purpose

Our primary hypotheses are:

(1) Precise characterization of COPD subjects using computed tomography – as well as clinical and physiological measures assessed longitudinally – will provide insight that will enable the broad COPD syndrome to be decomposed into clinically significant subtypes.

(2) Genome-wide association and DNA sequencing studies will identify genetic determinants for COPD susceptibility that will provide insight into clinically relevant COPD subtypes.

(3) Distinct genetic determinants influence the development of emphysema and airway disease.

Specific Aim 1: Cohort Building – Phase 1
(This Aim has been completed except for enrollment of additional non-smoking control subjects.)

Identify and phenotype up to 4,500 COPD cases GOLD Stages 2 through 4, 4,500 smokers without COPD, 1,500 GOLD Stage 1 and GOLD U subjects, and up to 1,500 non-smokers without lung disease, from two racial/ethnic groups (non-Hispanic Whites and African Americans) for genetic, epidemiologic, and natural history studies.

Specific Aim 2: Genome-Wide Association Study – Phase 1 and 2
(Analytic work is in progress.)

a. Stage 1. A genome-wide panel of SNPs will be tested for association with COPD in case-control samples from non-Hispanic Whites and African Americans.
b. Stage 2. Fine mapping of candidate genes to identify susceptibility alleles and/or high risk haplotypes using multiple study designs and independent samples, including:
    - Entire set of case-control samples from both racial/ethnic groups
    - External validation using family-based association analysis in the Boston Early-Onset COPD Study and the International COPD Genetics Network.
c. Stage 3 (Part of Phase 2 of COPDGene). We will genotype the Exome Chip (Illumina, Inc.) in all COPDGene smokers and test for rare and common variant associations with baseline and longitudinal COPD-related phenotypes. We will perform whole sequencing of subjects with distinct imaging characteristics to identify rare and common genetic variants influencing COPD susceptibility, emphysema, and airway disease.

Specific Aim 3: Characterization of Subtypes of COPD – Phase 1 and 2
(Analysis in progress.)

a. To further characterize the unique airway and parenchymal phenotypes of COPD by determining their associations with clinical, physiologic and functional indices.
b. Identify susceptibility genes for COPD subtypes, including CT-defined emphysema and CT-defined airway disease.

Specific Aim 4: Natural History of COPD and Risk Factors for Progression
The five-year natural history and progression will be completed in Phase 2.
a. Five years after the initial clinical evaluation, a second visit will be performed in all subjects enrolled in the study to assess disease status based on clinical, physiological, and chest CT assessments similar to the initial study visit.
b. The cohort will also continue to be followed longitudinally with telephone questionnaires every six months to determine mortality, COPD exacerbations, current cigarette smoking and co-morbid disease events.

Subject Selection

We are recruiting up to 12,000 smoking and non-smoking subjects.

Subject Inclusion/Exclusion Criteria for Phase 1 Data
A total of 12,000 non-Hispanic White and African-American subjects will be or have been recruited. Data from the Phase 1 study visit has been stored in the Data Coordinating Center. Biological specimens have been stored at the Johns Hopkins Biorepository and at the Brigham and Women's Hospital Biorepository. CT images have been stored in the National Jewish Imaging Core as well as the Brigham and Women's Hospital and University of Iowa Imaging Cores.

Phase 1 Actual Enrollment:
Phase 1, which was performed from November 2007 until July 2012, enrolled 10,364 subjects. Of these subjects 3416 were African American and 6884 were Non-Hispanic White. Enrolled subjects classified by severity of lung disease were as follows:

Smoker Controls – 4388
GOLD 2 - 4 Subjects – 3690
GOLD Unclassified – 1257
GOLD 1 Subjects – 794
Non-smokers controls – 108
Lack final GOLD classification due to failed spirometry – 63
Significant ILD noted on CT scan, excluded from analysis – 64

Inclusion Criteria for the COPDGene Central Study Protocol:
Enrollment in the COPDGene study at any of the twenty one clinical centers with subject identifiers and protected health information uploaded to the Data Coordinating Center.

Exclusion Criteria:
None, if the inclusion criterion is met

Enrollment of up to a total of 1700 additional non-smoker controls is anticipated in Phase 2.

Study Procedures Related to COPDGene Central Study Components

During the Phase 2 research study visit, the following procedures will be performed at the local Clinical Centers but are related to Central Study functions. Local Clinical Center functions are described to provide a context for the Central Study functions.

1. At the Clinical Centers: Informed consent will be obtained prior to any other study procedure. An updated HIPAA consent to obtain medical records to adjudicate the cause of death will be obtained. Subjects will be asked to provide a signed release of medical records to be used in the event of their death, to obtain a next of kin interview/ knowledgeable friend interview or hospital and physician records related to the events and illnesses associated with their death.
2. At the Clinical Centers: Subjects will be asked to provide updated contact information (address and two phone numbers) including two secondary contacts with addresses and phone numbers. The subjects will be asked to give consent to have their participation in the COPDGene study disclosed to these secondary contacts in the event that we are unable to contact them, and also disclosed to their next of kin, personal representative or

family members in the event of their death. Subjects will also be asked to permit transmittal of their personal information including social security number to the COPDGene Data Coordinating Center to use for central subject tracking, longitudinal follow-up, death searching through the Social Security Death Master File (SSDMF) and obtaining cost and utilization information from the Centers for Medicare and Medicaid Services (CMS) databases.

3. At the Clinical Centers: Contact information (but not social security number) will be collected from two other individuals likely to know the subject's whereabouts, at least one of whom is a relative not living with the subject. The purpose of this information is to locate subjects who have moved to a different home and have a different address and/or phone number and ascertain vital status of the subject.
4. Central Study Function: Phase 2 blood samples will be stored at the COPDGene Biological Repository at Brigham and Women's Hospital.
5. Central Study Function :CT Images will be transmitted to the COPDGene Imaging Cores at three locations for Analysis
6. Central Study Function : Data (Protected Health Information and Subject Identifiers) from COPDGene subjects will be transmitted to the Data Coordinating Center from the twenty-one Clinical Centers

Study Procedures Performed in Part by Both Clinical Centers and COPDGene Administrative Core

Longitudinal Follow-Up

Subjects will continue to be contacted up to four times per year by telephone, mail, or email for up to 10 years. Questions will be asked about current health status, exacerbations, cancer, new illnesses or medical conditions and current smoking status. The longitudinal follow-up contact mechanism is primarily based on automated contacts to subjects via a computer server controlled by the local clinical center in which the clinical center securely uploads subject contact data and social security number to a secure server using secure sockets (SSL) technology and 128 bit or greater encryption with an HTTPS protocol. Subjects establish a preference for contact by email and web data entry or automated phone calls with data entry by telephone keypad. Subject contact information and identifying information is deleted automatically from the server after the contact is made or at the end of three weeks. Data collected from the subjects is de-identified and made available to the Data Coordinating Center at National Jewish to be stored. Subjects who fail to respond to automated contacts are contacted by the local clinical coordinator who then completes the clinical questions for the subject into the web-based data collection form. Subjects who become lost to follow-up from the longitudinal follow-up process are traced by their secondary contacts and searched for in the social security death index using the same automated server system querying the social security master death file. Deaths identified from the social security death search are communicated to the Data Coordinating Center and the local clinical center. The local clinical center initiates collecting further information about the death as described below.

Central Study Function: In the event that the local clinical center is unable to complete the longitudinal follow process for their subjects the COPDGene central Administrative Core will take over that function using the stored subject personal identifiers to upload subject information to the computer server and will complete follow-up coordinator phone calls to subjects.

Communications to Study Subjects
In addition to the longitudinal follow-up program of phone and email contacts, Subjects may be contacted on no more than three additional occasions per year to inform them of other research studies and to update them about results of the COPDGene study. In general these contacts will be made through staff at the local clinical center.

Central Study Function: However selected communications may be made by mail through the COPDGene Administrative Core for efficiency or cost reasons after there has been local or central IRB approval of the content of the communication.

Central Study Function: Because the COPDGene Cohort is a national resource for understanding the impact of smoking on human disease and particularly on the lung, it is essential to retain the subjects effectively for the duration of possible research funding. Local clinical centers may not retain staff during periods when no active study visits are occurring, but maintaining contact and longitudinal data collection is necessary. The central Administrative Core of the study can assume subject follow-up contacts when local clinical centers are unable to complete this work. For clinical centers that are unable to complete study mailings or phone contacts, those functions can be assumed by the COPDGene Administrative Core when subject identifiers have been provided to the DCC.

Transfer of Subject Identifiers to the Phase 2 Data Coordinating Center

The process of transferring subject identifiers to the Phase 2 DCC will be accomplished in several steps and will be dependent on local approval at the clinical centers for the process.

1. For subjects who enrolled in Phase 1 of COPDGene and whose HIPAA compliant authorization or informed consent included explicit authorization to transfer subject identifiers to the DCC, we will request that data be sent as soon as the Phase 2 DCC is approved to receive the data and has completed data security preparations as described in this protocol.

2. For subjects whose Phase 1 consent and HIPAA authorization are ambiguous regarding transfer of identifiers we will ask the local centers to request permission to transfer the data centrally to the DCC under the supervision of their local IRB.

3. In the rare cases where clinical centers and local IRB do not approve transfer of the subject identifiers from Phase 1 we will have the option to send a mail consent to the subjects involved and request explicit permission to store identifiers in the DCC.

4. For all Phase 2 subjects we will include transfer of identifiers in the protocol, consent and authorization documents.

Mail Consent procedure:
Central Study Function: The central study will provide a consent addendum/HIPAA authorization template and a cover letter to the local Clinical Centers to be mailed to Phase 1 subjects who have not returned for a Phase 2 visit. In the mailing, two copies of the informed consent addendum should be included. The cover letter instructs the subject to review the informed consent addendum and sign both copies if they are willing to participate. They should return one signed copy in the enclosed pre-addressed envelope. In addition a memo will be included that the subjects can give to their next of kin or personal representative explaining their participation in the study and planned follow-up with them in the event of death. The mailing will

also include a form to provide updated contact information for the subject, their secondary contacts and their current treating physician.

The purpose of the informed consent addendum is to invite the subject to consent to an improved process to confirm vital status and obtain information about cause of death in the subjects who die before they are able to participate in the phase 2 visit. The consent addendum will allow the subject to designate next of kin, friends or a personal representative who can be contacted in the event of their death and allow that individual to disclose information about events surrounding their death. It will also allow the subject to authorize release of records from treating physicians and hospitals that they are admitted to. It will explicitly allow research data collection. and will also include a provision to transfer subject name, address, social security number and phone numbers to the Data Coordinating Center and Administrative Core to be used for central data collection, including death searching the social security death masterfile, collecting records related to death and obtaining Medicare cost and usage information. The subject identifiers will be handled and stored in the same fashion as described for the Phase 2 subject visit data transfer.

If subjects cannot be reached or do not agree to sign the addendum, there will be no change in their status within the study. They will be contacted as usual for their Phase 2 visit and will continue having periodic contacts by email or telephone for updates on their health through the LFU system. Local clinical centers will continue to track subjects as described in the Phase 1 consent without a specific HIPAA authorization for medical records after death. Information regarding whether the subjects have signed the consent addendum will be maintained in the subject tracking system that is available to the local clinical center coordinators and the central administrative core.

Mortality Assessment

Local Clinical Center and Central Study Function: Death is an endpoint of interest that will be analyzed as part of this protocol. The opportunity to assess vital status will occur in one of several ways. It is possible that during the process of longitudinal follow up, the clinical center will be made aware of a subject's death. However, some subjects will be lost to follow up and vital status unknown. For these subjects all clinical centers will conduct a search to determine vital status. This search will include periodic searches of the Social Security Death Master File and general internet search including obituary postings. Once the individual's vital status has been confirmed as deceased, the clinical center will obtain additional information regarding the subject's death so that cause of death may be determined. This additional information will include: 1) Informant Interview conducted with next of kin or other close contact of the study participant or the subject's physician, 2) Death certificate, 3) hospital Discharge Summary if it is determined that the death occurred in hospital or within 28 days of a hospitalization, and 4) request for recent treating physician records if death did not occur in the hospital.

Central Study Function: The above records will be de-identified and transferred to the Data Coordinating Center. These records will be centrally reviewed by a Mortality Adjudication Committee. Cause of death will be assigned to one of the following causes: respiratory, cardiovascular, cancer, other. The data will be reviewed by at least two members of the Adjudication Committee not associated with the institution where the death occurred and a final cause of death assigned according to the Principles of Adjudication. If the two committee members disagree on cause of death, a third member of the committee will review death records and cause of death will be assigned based on majority consensus.

Death adjudication will be coordinated by the central Administrative Core and all materials (death certificates, hospital records and physician records) for death adjudication will be maintained centrally. Death searching through the social security death master file will be automated through the longitudinal data collection server when permitted by local IRBs and manual searching will be performed by the central Administrative Core for subjects who have centralized personal information. Information on cost of care will be obtained from Medicare Databases by the central Administrative Core using the social security numbers stored centrally.

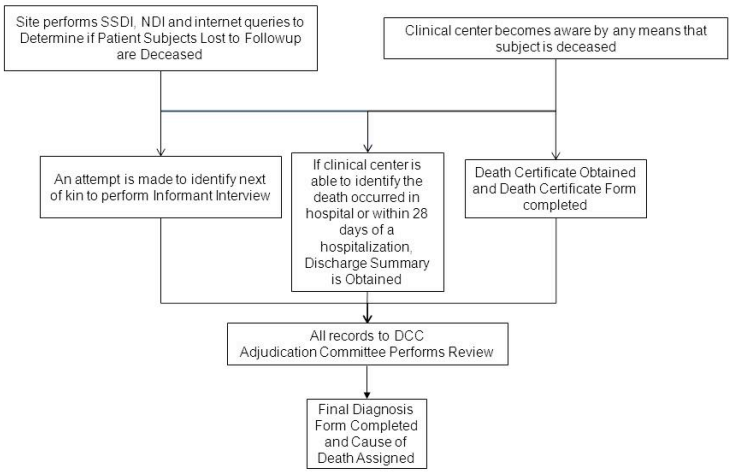

Figure 2 Procedures and Flow Diagram for Death Tracking and Death Adjudication

Study Procedures Performed by Imaging Core

The Imaging Core in its three locations (National Jewish Health, Brigham and Women's Hospital and University of Iowa) receives CT images from the twenty one clinical centers. The images are accompanied by the date of study which is essential to the analysis in order to compare scans from Visit 1 and Visit 2. These images are analyzed for quantitative variables of emphysema, gas trapping and others. The images are also visually scored by investigators. Images can be fully de-identified when research procedures do not require knowing the date of study in order to be shared as a de-identified dataset. When images are shared outside of the Imaging Core with the date of study intact, the files will be treated as a limited dataset.

Study Procedures Performed by PFT Core

The PFT Core in two locations (National Jewish Health and Dr Robert Jensen as a contract consultant) will receive protected health data from the spirometry and DLCO testing and the date of the study. After completing the quality review of the Spirometry and DLCO the data will be returned to a file depot in the DCC. It will be added to the main database. For Spirometry and DLCO results that fail quality assurance, a secondary review will be made by the PFT committee.

Data Collected and Data Handling

Each Clinical Center involved in the study will follow the COPDGene® Study protocol already approved by their local IRB to obtain data for the study. Any information collected from any subject involved with COPDGene® will be given a subject ID that consists of letters, numbers, and a three letter center code

Phase 1 Data
The data collection for Phase 1 has been done through the COPDGene Study under local IRB approval for the project. This data is currently stored in the Data Coordinating Center as *existing* de-identified data that is described below.

Questionnaires: Questionnaires were administered to collect information on lung and selected medical conditions, lung surgery, smoking history, previous studies, recent surgery and hospitalizations, heart conditions, and age, race and ethnicity. Following the initial study visit, subjects are contacted every 6 months to complete a longitudinal follow up questionnaire. All questionnaire data to date has been stored at the DCC.

Breathing Test (Spirometry): Spirometry was collected on all subjects. Both a pre-bronchodilator and a post-bronchodilator measurement are collected from each subject.

Physical Assessment: Subject blood pressure, height and weight were obtained.

Six-Minute Walk Test: Subjects were asked to walk for 6 minutes on a level surface following a standard protocol, and a measurement of total distance walked was obtained.

High Resolution Chest CT Scan: Chest CT scans were obtained for all subjects following the COPDGene® CT scan protocol. These scans were read by the clinical center radiologist and then de-identified and transferred to the COPDGene® Imaging Core. Output from the analysis of the CT scans is stored in the DCC as de-identified data.

Exacerbations and Incident Medical Conditions: Subjects are contacted twice a year using a combination of automated phone calls, email and coordinator phone calls. Data output from these contacts are stored in the DCC as de-identified data.

Subject Vital Status: Information about deaths in the cohort under Phase 1 informed consent of the study is obtained from publicly available sources including unsolicited reports from family or friends, published obituaries, failure to respond to subject contacts and searching the social security death master file.

Phase 2 Data
The data collection for Phase 2 will be done through the COPDGene Study under local IRB approval for the project. Subjects will provide new informed consent to participate in Phase 2. Data will be transferred within the study cores as shown in Figure 1. The DCC will receive and store identifiable data after receiving IRB approval for this activity.

Data Collection will include:

Study Questionnaires: (Respiratory Symptoms, Medical History, Medications, SF-36, St George's, HADS, Longitudinal follow-up questionnaire, Residential and Environmental, CAT, For Women Only, and Socioeconomic).

Physical Assessment (modified in phase 2 to include arm span and standardized blood pressure)

CBC Results: Recorded on data form uploaded to DCC

Spirometry: Results with date of test

DLCO (new in phase 2): Electronic files transferred directly to DCC without personal identifiers but with date of study

Six Minute Walk Test: results recorded on data form and uploaded to DCC

CT Scan of the Chest: image files that are de-identified of subject identifiers except for date of study

Personal identifiers: Name, address, date of birth, social security number, names and addresses of secondary contacts, date of death when applicable.

Death Adjudication: Data will be collected regarding cause of death to include death certificate, next of kin interview, physician interview and office records and hospital records. Records may include identifying information such as name, address and social security number when subjects have provided consent for identifying information to be transmitted to the central study. Records will be redacted of identifying information by the local clinical center prior to transmission if permission has not been granted.

Data Storage and Distribution

The Division of Biostatistics and Bioinformatics at National Jewish Health (NJH) is the Data Coordinating Center (DCC) for COPDGene®. The COPDGene® DCC will maintain the confidentiality of all protected health information collected under this protocol using physical security, database security, and web applications security.

Physical Security

Computer Room. The Division of Biostatistics and Bioinformatics maintains its own computer server room which is on the same floor as the Division offices; access to this server room is tightly restricted. This server room has dedicated power, cooling, lighting, and an environmental monitoring system that alerts—by pager—two on-call systems engineers when temperature or humidity levels rise above acceptable limits. By January 2014 the Division will move its servers into the enhanced IST computer server room; this will be an even more secure environment than we maintain at present.

Backup. Each server and workstation in the Division is backed up to tape. Daily and weekly backups are stored on-site in a fire safe for easy access and quick recovery. Weekly, monthly, and quarterly backups are archived off-site. A total of 2 years' backups are maintained in storage. In order to ensure confidentiality, all patient-related data are encrypted during tape backup; for emergency decryption only, the encryption keys are recorded on a flash drive which is stored in a locked safe.

Database Security
The main COPDGene® database is stored in a Microsoft SQL 2005 database on a server dedicated to Microsoft SQL Server 2005 (SQL server). Each SQL database defines the users who may access the database. The COPDGene® DCC incorporates Windows Authentication to define the members of *user groups* who can access a specific SQL database. In addition, we define permissable actions for each *user group* using Microsoft SQL's *grant* and *deny* commands.

The main COPDGene® database uses a deidentified subject identification code to associate a particular subject with data submitted on a data collection form or data collected using a medical device such as a spirometer or a CT scanner. A subject's date of birth is collected on a demographics form and is used solely to calculate age; date of birth is not included in the main COPDGene® database that is provided to investigators. When data from a spirometer or a CT scanner are first submitted to the DCC, however, the spirometer data and the CT data do include the date of service.

In the second phase of COPDGene®, a new Subject Personal Information form will collect elements of Subject Identifiers, including name, address, and social security number. The data from this form will be used by the Administrative Core of COPDGene® for death adjudication, but all the subject identifiers will be stored in a SQL database that is separate from the main COPDGene® database; moreover, the SQL database with subject identifiers will reside on its own SQL server. Microsoft SQL has its own encrypted backup system, and all subject identifier data will be backed up in an encrypted form. Direct access to this COPDGene® subject identifier database will be restricted to essential DCC personnel; it will have a distinct owner, and it will not use the default SA owner; last, the *guest* user has been removed from this server. These measures reduce even further the chances that someone could hack their way into this subject identifier database.

Two members of the Administrative Core, Associate Director, Dr. Elizabeth Regan and the COPDGene® Central Study Project Manager, will have access to a single encrypted copy of the subject identifier database. This copy will reside on a single desktop computer in Goodman K706 and will be encrypted using TrueCrypt (http://www.truecrypt.org/), free open-source disk encryption software that is currently being used within NJH for Honest Broker protocols associated with the NJH Research Database. At this time, it is not anticipated that it will be necessary to store an additional encrypted copy of this subject identifier database on a portable electronic device.

The only elements of the subject identifier database that we expect to associate with protected health information (PHI) are addresses and occupations in order to examine the relationships of environmental and occupational history to the development and progression of COPD. Even then, we expect that these limited datasets containing addresses will be distributed only to select Principal Investigators at participating Centers; these Investigators will apply to obtain these data and will receive a copy of the database only if their application is approved by the Executive Committee and a Data Use Agreement is in place.

Web Applications Security
The DCC maintains a password-protected, limited-access COPDGene® website that is located at a secure https URL. By definition, https websites are secured using Secure Sockets Layer (SSL) or Transport Layer Security (TLS) technology standards. SSL is used to encrypt data transmissions between the web server and a user's computer. In addition, the DCC maintains a security certificate that authenticates the website resides within NJH.

Among its other functions, the COPDGene® website provides the ability to submit new data and edit existing data. Access to these features of the COPDGene® website is controlled by defining distinct user groups that have different capabilities. The DCC will review on a monthly basis login activity from the Clinical Centers that participate in COPDGene®. If a user has not logged in to the COPDGene® within the previous 30 days, that's user's access to the web site will be suspended until their status with COPDGene® can be confirmed.

Although study site coordinators can review data for subjects from their Clinical Center, they cannot make direct changes to their data: if they discover an errant datum, they must request a change to that datum. This request is logged in an audit log and acted upon by DCC personnel. Because PHI data in COPDGene® will reside on a server separate from the main COPDGene® database, the DCC will build a custom interface through which study coordinators can review and request edits to PHI data.

Distribution of Data and Biospecimens

The COPDGene® Central Study acting through established policies in the Administrative Core and the Ancillary Studies and Executive committees will oversee and coordinate the distribution agreements for collaborating investigators through the Data Coordinating Center (DCC). Tracking of those distribution agreements will be managed by the Administrative Core staff. This includes verifying that Collaborating Investigators and the Recipients sign the appropriate agreements, as well as keeping those documents on file and reviewing data requests for appropriateness and merit. They are responsible for ensuring data and images from COPDGene are stored securely and managed in accordance with existing regulations for protection of privacy and human subject research.

There are three major types of data distribution that are anticipated from COPDGene Study: 1) Data sharing for "public" access to NIH funded data via dbGaP, 2) Data distribution to investigators within the COPDGene study and 3) Data distribution to investigators outside of the COPDGene study who have approved ancillary studies.

1. Data Sharing to dbGaP

The COPDGene Data Coordinating Center (DCC) has acted to oversee transfer of *existing* de-identified subject data to the NIH controlled dbGaP from Phase 1 of the study. dbGaP provides storage, oversight, and a process for distribution of phenotypic and genetic data as required by the NIH Data Sharing agreement. The DCC will oversee future data transfers of de-identified data to dbGaP. dbGaP is responsible for review and oversight of all data released from their database. Any request for data from researchers who are not associated with the COPDGene study and do not have approved ancillary studies will be directed to dbGaP.

2. Data Sharing with Internal COPDGene Investigators

Internal COPDGene Investigators are defined by the Executive Committee to include clinical center directors and co-investigators, members of all study cores and data analysis working groups, investigators associated with the Industry Advisory Committee and other interested investigators who have been invited to join in the study.

<u>De-identified Datasets</u>

The DCC will periodically post <u>de-identified</u> datasets at the instruction of the Executive Committee on the COPDGene password protected website, for the use of internal investigators. Internal investigators consist of co-investigators at the twenty one clinical centers and at the

study cores. These members of the study have been provided logins/passwords to access the internal COPDGene website. Internal investigators are advised that the datasets downloaded from the COPDGene website are provided for their individual use and may be shared within their local working groups that may include other research collaborators working with them and statisticians who are performing analytic work. Release of the de-identified dataset outside of the local research working group is not permitted without explicit permission of the COPDGene Executive Committee.

Limited Datasets and Biospecimens

CT images and Biospecimen requests may involve additional procedures before being released to internal or external investigators. The COPDGene Study has an obligation to ensure that these are released in compliance with the subject consents and privacy regulations. CT images are stored with the date of image acquisition in order to permit proper sequential analysis. Thus these image files represent "limited datasets" under HIPAA regulations. Investigators who request access to the CT images may ask that they be fully de-identified if data of study is not needed, or may submit a request for release of a limited dataset.

For requests from investigators associated with the COPDGene study, written proposals for acquisition of specimens, images with date, and clinical data with any subject identifier are initially reviewed by the COPDGene® Executive Committee. In the case of materials, the proposal will be discussed by the COPDGene administrative staff with the investigator to resolve any questions or missing information. During this time, the statistical plans and power analysis will be reviewed. Once the proposal is finalized, it is sent to the COPDGene® Executive Committee who will then review each proposal for merit and feasibility. For co-investigators who are requesting a limited dataset that may include any subject identifiers, they will be required to have local IRB approval of their project and provide the COPDGene study with a plan for securing the limited dataset and its ultimate disposition. If these individuals are not located at institutions that have an existing contract with the COPDGene study, they will need to establish a data use agreement with National Jewish to receive limited datasets or specimens. Upon approval, the COPDGene® Executive Committee will instruct the Biorepository, QIL or DCC to release the approved data, specimens or images.

Because storage of biological samples from research subjects are subject to the storage and sharing constraints imposed in the subject consent form, biospecimens that are provided to investigators must either be returned to the COPDGene Biorepository, destroyed, or used completely; and the receiving investigator must agree to the conditions and comply/ attest to the final disposition. Biospecimens cannot be transferred to another investigator or used for other analyses without approval of the COPDGene Executive Committee.

3. Data Sharing for Investigators Outside of COPDGene and their Ancillary Studies

In general, ancillary studies and associated requests for data, access to research subject or biospecimens will be made to the ancillary study committee using the Ancillary Study Proposal Form (Appendix V).

De-identified Datasets

De-identified datasets can be provided to investigators outside of COPDGene with or without an approved ancillary study, at the discretion of the COPDGene Executive Committee. Investigators can also be referred to dbGaP for access to the de-identified data. An ancillary study proposal with collaborators at Pfizer will use de-identified data (genetic, imaging and phenotypic) structured as a data analysis contest and posted publically on the internet to analyze complex associations that may provide additional insight into COPD pathogenesis –

described as “Crowdsourcing”. The details of the proposal are included in an addendum to the Central Study Protocol (Appendix VI).

Limited Datasets and Biospecimens

For requests from investigators outside the COPDGene study, written proposals for acquisition of specimens, images with date, and clinical data with any subject identifier are initially reviewed by the COPDGene® Executive Committee. Limited Datasets and Biospecimen requests from investigators outside the COPDGene study must be accompanied by an approved IRB protocol. A data use agreement must be signed before any data or specimens can be released. The project must be reviewed by the Ancillary Study Committee and the Executive Committee. Before the specimens or data are released to the investigator appropriate portions of the Appendices I-II must be completed and signed.

As described for internal investigators, non- COPDGene investigators requesting biospecimens and images, will have their proposal reviewed and discussed by the COPDGene administrative staff. The staff will contact the investigator to resolve any questions or missing information. During this time, the statistical plans and power analysis will be reviewed. Once the details of the proposal are finalized by the administrative staff, it will be sent to the COPDGene® Executive Committee who will then review each proposal for merit and feasibility. For co-investigators who are requesting a limited dataset that may include any subject identifiers, they will be required to have local IRB approval of their project and provide the COPDGene study with a plan for securing the limited dataset and its ultimate disposition. Upon approval, the COPDGene® Executive Committee will instruct the Biorepository, QIL or DCC to release the approved data, specimens or images.

Because storage of biological samples from research subjects are subject to the storage and sharing constraints imposed in the subject consent form, biospecimens that are provided to investigators must either be returned to the COPDGene Biorepository, destroyed, or used completely; and the receiving investigator must agree to the conditions and comply/ attest to the final disposition. Biospecimens cannot be transferred to another investigator or used for other analyses without approval of the COPDGene Executive Committee.

Distribution of Biospecimens Overview

To begin a specimen/data request, investigators should submit an application using the format in Appendix II.. The COPDGene® Executive Committee will review the application and communicate its decision to the requesting investigator.

To protect the confidentiality and privacy of participants, Recipients of COPDGene Data and/or Biospecimens must adhere to the requirements of the appropriate Investigator Certification. Failure to comply with this Agreement could result in denial of further access to COPDGene® Data and Biospecimens. Violation of the confidentiality requirements of this agreement may leave requesting investigators liable to legal action on the part of COPDGene® participants, their families, or the U.S. government.

1. Application for Materials

An investigator wishing to use COPDGene® specimens along with images or other phenotypic data should submit an application or Ancillary Study Proposal to the COPDGene® Ancillary Study Committee. The application should include the title, investigators, hypotheses to be tested, identification of variables, specimens or images, analysis plan and proposed timetable.

2. Responsibility of Investigators

The investigator must agree to not distribute materials, to avoid the waste of precious materials, and to the return of materials when appropriate. Recipient will agree to retain control over Data, Genetic Analysis Data, and Biological Materials, their progeny, and unmodified or modified derivatives thereof, and further agrees not to transfer Data, Genetic Analysis Data or Biological Materials, their progeny, or unmodified or modified derivatives thereof, to any other entity or individual. The recipient will agree to make reasonable efforts to avoid contamination or waste of the samples during material handling.

Investigators must fulfill the following requirements:

a) Certify the investigators willingness to adhere to COPDGene® guidelines for confidentiality, use of specimens, biosafety and indemnification.

b) Before receiving COPDGene® data, specimens or images the investigator must indicate that the data, specimens or images will be used only as agreed upon in the collaboration and will document this at completion of the study.

c) Provide an annual progress report.

3. Credit and Authorship

All presentations and manuscripts shall acknowledge that the data were collected through COPDGene® (Appendix IV). If there is collaboration with COPDGene® researchers, authorship would be expected. The "right" to authorship is determined by substantive intellectual contribution of an individual to at least 2 of 3 areas: study question concept; data collection and analyses; and manuscript preparation. Appropriateness of authorship for a given study may be reviewed by the COPDGene® DSR.

Biostatistical Analysis

In Phase 1, genome-wide association analysis will be performed using genome-wide SNP genotyping data that has been obtained in the entire COPDGene study population. In Phase 2, exome chip genotyping, whole genome sequencing, and targeted sequencing data will be obtained. For exome chip and targeted sequencing data, analyses will be performed on a gene level, while for whole genome sequencing, sliding windows of 20 SNPs will be used. For rare variant analysis, we plan to use three analytical approaches to analyze non-synonymous SNPs—the Combined Multivariate and Collapsing Method (CMC) (14), the Morris and Zeggini (M-Z) likelihood ratio method (15), and a powerful weighted, permutation-based approach (16). For all three of these approaches, analyses will be performed at the gene level, using only nonsynonymous variants with allele frequencies < 0.05. We recognize that selection of this 5% threshold to define a rare SNP is arbitrary; therefore, we will also perform exploratory analyses using a lower allele frequency threshold of 1% as well as an analysis of all detected SNPs, regardless of allele frequency. Due to differences in genetic background, separate analyses will be performed in NHW and AA subjects.

In addition to rare variants, Exome Chip genotyping and whole genome and targeted sequencing will also identify common non-synonymous and other functional variants. Single common variant associations to baseline and longitudinal COPD-related phenotypes will be assessed using regression analysis in PLINK, separately in non-Hispanic Whites and African Americans, with adjustment for age, gender, pack-years, and principal components of global genetic ancestry (17). Although our primary focus will be on single SNPs, copy number variants will also be assessed for association to COPD and COPD-related phenotypes.

To provide conservative power estimates for our rare variant analysis, we performed a simulation study where DNA sequencing data were generated with the FREGENE package (18), which creates simulated data for realistic specifications of selection, recombination, migration and population structure. We simulated case-control status with disease prevalence set at 10%. With our proposed two-stage strategy of whole genome sequencing in 1334 subjects (corresponding to 667 emphysema or airway-predominant cases vs.667 controls) and follow-up sequencing in 8000 subjects, we randomly selected a sliding window region with 20 rare variants (i.e., MAF < 0.01). Affection status was independently caused by a varying percentage of disease susceptibility loci (DSL) in the 20 regional rare SNPs within a sliding window. The disease model assumes that the odds of disease are additive for each risk allele on the log scale. To avoid overestimating statistical power, the association between the rare variants and affection status is assessed based on the Ionita-Laza approach across the two stages. **Table 1** shows the estimated power for a variety of scenarios with different odds ratios and varying number of DSLs (1-20 DSLs out of 20 rare variants) in the entire study population. All estimates are based on 1000 replicates. With odds ratios of 2.25, power will be excellent, even if there is substantial allelic heterogeneity. Rare variant analysis using the Exome Chip in the entire study population will have even greater power, with 96% power to detect an odds ratio of 2.0 with 60% of the SNPs in the region as DSLs.

Table 1: Power to Detect Rare Variant Associations within COPDGene

	% of SNPs in Region That Are Disease Susceptibility Loci				
Odd Ratio for Disease Susceptibility Locus	**100**	**80**	**60**	**40**	**20**
2.5	100	100	100	100	100
2.25	100	100	98	95	78
2.0	99	94	79	32	9

Risks and Discomforts

For the research procedures that are described in this protocol the major risk associated with the study is inadvertent breach of confidentiality with disclosure of health information linked to subject identifiers. With storage of social security numbers there is also the potential for financial impacts and identity theft if social security numbers are inadvertently lost or disclosed.

We will be storing genetic data in addition to subject identifiers. An inadvertent breach of confidentiality of this data may affect the subject's employment or ability to obtain health care. A Certificate of Confidentiality has been obtained for the COPDGene study to provide additional protection for study participants in relationship to the genetic data that we will be holding.

This study is designed to be a national resource for scientific investigations. As such, medical information, genetic information and samples will be provided to dbGaP in order to make the study data available other investigators, with appropriate safeguards. Other researchers interested in using such information for scientific investigations will be required to apply to the Executive Committee for permission to access the data for studies that have received local IRB approval and with requirements to maintain subject confidentiality.

This study is designed to be a national resource for scientific investigations. As such, medical information, genetic information and samples will be provided to dbGaP in order to make the

study data available other investigators, with appropriate safeguards. Other researchers interested in using such information for scientific investigations will be required to apply to the Executive Committee for permission to access the data for studies that have received local IRB approval and with requirements to maintain subject confidentiality. Subject identifying information will not be transmitted to other investigators. A Certificate of Confidentiality has been obtained for the COPDGene study to provide additional protection for study participants.

Potential Benefits

There are no expected benefits to the study participants. Improved understanding of COPD has occurred; COPDGene has generated over 45 scientific publications. Chest CT scans could identify pulmonary nodules, early lung cancer or other abnormalities that may require follow-up outside of this study.

Monitoring and Quality Assurance

This is an observational longitudinal investigation without a therapeutic intervention. It is expected that there will be deaths in both control and COPD subjects enrolled in this study that are not related to study procedures. It is expected that there will be hospitalizations from a variety of causes not related to study procedures including but not limited to newly discovered disorders, acute disorders requiring surgery, pre-existing conditions, and exacerbations of underlying COPD. Subjects may expire due to pre-existing or new diseases including cancer, cardiovascular conditions and COPD. These are anticipated events that are not related to this investigation. These events will not be prospectively collected as part of the current study and thus will not be reported to IRBs. There are no expected Serious Adverse Events in this study related to study procedures. At each Clinical Center, subjects will be observed for the development of tremulousness, and nervousness following bronchodilator medication. Unexpected Adverse Events related to study procedures will be reported to the IRB of the Clinical Center and to the Executive Committee. An Observational Safety and Monitoring Board (OSMB) has been appointed by the National Heart, Lung, and Blood Institute and will continue to oversee this study.

Quality assurance of spirometry data will be insured by the Pulmonary Function Core, which will review spirometry data from each study participant. Quality assurance of CT scans will be analyzed by the Imaging Core in Denver. Questionnaires and other data will be quality controlled by the Data Coordinating Center in Denver.

As noted above we anticipate that some subjects may expire during the next phase of this study due to a combination of pre-existing disease and the onset of new conditions. These events are not anticipated to be related to the study visit; however, they provide important information about the natural history of COPD and other smoking-related conditions. We will monitor and collect information about deaths in the cohort but will not report them to IRBs as study-related events unless they occur during a study visit or within twenty-four hours of the study visit and are judged to be related to a study procedure.

References

1. Hoyert DL, Xu J. Deaths: Preliminary Data for 2011. In: U.S. Department of Health and Human Services. Centers for Disease Control and Prevention NCfHS, National Vital Statistics System, editor. 2012. p. 65.
2. Burrows B, Knudson RJ, Cline MG, Lebowitz MD. Quantitative relationships between cigarette smoking and ventilatory function. Am Rev Respir Dis. 1977; 115:195-205.
3. Silverman EK, Chapman HA, Drazen JM, Weiss ST, Rosner B, Campbell EJ, et al. 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. 1998; 157:1770-8.
4. Silverman EK, Palmer LJ, Mosley JD, Barth M, Senter JM, Brown A, et al. Genomewide linkage analysis of quantitative spirometric phenotypes in severe early-onset chronic obstructive pulmonary disease. Am J Hum Genet. 2002; 70(5):1229-39. PubMed PMID: 11914989.
5. Hersh CP, DeMeo D, Silverman EK. COPD. In: Silverman EK, Shapiro SD, Lomas DA, Weiss ST, editors. Respiratory Genetics. London: Hodder Arnold; 2005.
6. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet. 2005; 6(2):95-108. PubMed PMID: 15716906.
7. Global Initiative for Obstructive Lung Disease. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease (Updated 2011) December 2, 2012. Available from: http://www.goldcopd.com.
8. Vestbo J, Hurd SS, Agusti AG, Jones PW, Vogelmeier C, Anzueto A, et al. Global Strategy for the Diagnosis, Management and Prevention of Chronic Obstructive Pulmonary Disease, GOLD Executive Summary. Am J Respir Crit Care Med. 2012. Epub 2012/08/11. doi: 10.1164/rccm.201204-0596PP. PubMed PMID: 22878278.
9. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. Am J Respir Crit Care Med. 1999; 159(1):179-87.
10. Ferris BG. Epidemiology Standardization Project. Am Rev Respir Dis. 1978; 118 (suppl.):1-120.
11. Society AT. ATS statement: guidelines for the six-minute walk test. Am J Respir Crit Care Med. 2002; 166(1):111-7. PubMed PMID: 12091180.
12. Macintyre N, Crapo RO, Viegi G, Johnson DC, van der Grinten CP, Brusasco V, et al. Standardisation of the single-breath determination of carbon monoxide uptake in the lung. Eur Respir J: official journal of the European Society for Clinical Respiratory Physiology. 2005; 26(4):720-35. PubMed PMID: 16204605.
13. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand. 1983; 67(6):361-70. PubMed PMID: 6880820.
14. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. Am J Hum Genet. 2008; 83(3):311-21. Epub 2008/08/12. doi: 10.1016/j.ajhg.2008.06.024. PubMed PMID: 18691683; PubMed Central PMCID: PMC2842185.

15. Morris AP, Zeggini E. An evaluation of statistical approaches to rare variant analysis in genetic association studies. Genet Epidemiol. 2010; 34(2):188-93. Epub 2009/10/08. doi: 10.1002/gepi.20450. PubMed PMID: 19810025; PubMed Central PMCID: PMC2962811.

16. Ionita-Laza I, Buxbaum JD, Laird NM, Lange C. A new testing strategy to identify rare variants with either risk or protective effect on disease. PLoS Genet. 2011; 7(2):e1001289. Epub 2011/02/10. doi: 10.1371/journal.pgen.1001289. PubMed PMID: 21304886; PubMed Central PMCID: PMC3033379.

17. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38(8):904-9. Epub 2006/07/25. doi: 10.1038/ng1847. PubMed PMID: 16862161.

18. Chadeau-Hyam M, Hoggart CJ, O'Reilly PF, Whittaker JC, De Iorio M, Balding DJ. Fregene: simulation of realistic sequence-level data in populations and ascertained samples. BMC Bioinformatics. 2008; 9:364. Epub 2008/09/10. doi: 10.1186/1471-2105-9-364. PubMed PMID: 18778480; PubMed Central PMCID: PMC2542380.

Principal Investigators
JAMES D. CRAPO, MD
National Jewish Health
EDWIN K SILVERMAN, MD, PHD
Brigham and Women's Hospital

Appendix I. Data Sharing and Access Policy

September 15, 2014

Genetic, Imaging and Phenotype Data on the COPDGene Cohort

Access to and use of the COPDGene data requires protection of the confidentiality of the subjects and assurance of scientific integrity with regard to the use of the data. COPDGene data can be accessed through dbGaP and also can be directly accessed through collaboration with COPDGene investigators. An institution, company or investigator who seeks access to the COPDGene data directly from the COPDGene Investigators will be required to agree to the following:

1. All data (genetic, imaging and phenotype) obtained from COPDGene will be used in a manner that protects the privacy and confidentiality of the participant subjects. No attempt will be made to identify subjects or link data to personal information that is readily identifiable or to contact COPDGene subjects in relationship to genetic, imaging or phenotype data.
2. Analysis of the COPDGene data will be limited to conformance with the consent groups for genetic studies. Some COPDGene subjects have agreed to allow use of their genetic information into any disease, while others limited their sample use to smoking-related disorders.
3. The data will not be distributed to other individuals without approval from COPDGene.
4. The data will be treated confidentially, physically secured, and will not be directly accessible from the internet. Authorized users will adhere to information technology practices in all aspects of data management to assure that only authorized individuals can gain access to the COPDGene data sets.
5. The data must be destroyed if requested by the COPDGene Executive Committee. It is understood that this applies to the original COPDGene data; it does not apply to new, appropriately de-identified data that may be generated.
6. The COPDGene Executive Committee will be notified within twenty-four (24) hours of the user becoming aware that there has been any unauthorized data sharing, breaches of data security, or inadvertent data releases that may compromise data confidentiality.
7. Manuscripts resulting from COPDGene data must be submitted for approval by the COPDGene Executive Committee prior to publication or presentation at public meetings.
8. The COPDGene data, in whole or in part, may not be sold to any individual at any point in time for any purpose.
9. An investigator may not move the COPDGene data from one institution or company to another, in whole or in part, without the written approval of the COPDGene Executive Committee.

10. Use of COPDGene data on portable media, such as a CD, flash drive, or laptop, is discouraged. A limited data set (containing subject identifiers) must be encrypted when stored on a portable device.

Data Request Form

Send Data Request Form to: Elizabeth Regan, MD, PhD, Associate Director of COPDGene: ReganE@NJHealth.org

* The costs associated with the transfer of data are the responsibility of the requestor. Cost estimates can be provided prior to submission of the proposal.

Investigator: _____
Date: _____
Title of Ancillary Study: _____
Ancillary Study Number: _____

☐ Smoking-related disease study ☐ Other Disease(s) not related to smoking

Subject Selection (check one or more)

☐ Final Analysis Primary 10,300
☐ Exclusionary Disease: ILD & Bronchiectasis (n=64)
Or Specific GOLD classification
☐ GOLD 0 – Smoker Controls (n=4388)
☐ GOLD 1 (n=794)
☐ GOLD 2 (n=1922)
☐ GOLD 3 (n=1162)
☐ GOLD 4 (n=606)
☐ GOLD Unclassified (n=1257)
☐ Nonsmoker Normals (n=108)

☐ Match predefined subject list (attach .csv, .xls, etc.)

☐ Custom criteria (ex. "if percent Emphysema > 40%") Specify in Additional Comments.

Output File Format

☐ SAS ☐ JMP ☐ Tab-delimited TXT

Phenotypic Data

☐ Visit 1 Phenotype data

Genetic Data

☐ COPDGene GWAS, PLINK format (9970 subjects, file size= 1.5GB)
☐ GWAS with Imputed Data, PLINK format (9970 subjects, file size=138.5 GB)
☐ Exome Chip (2313 subjects, file size=145 MB)
☐ Exome Sequencing

To reduce file size on GWAS or Imputed data, you can request specific SNPs and/or chromosomes:
Specific Chromosome _____
Specific SNPs and/or region _____

Imaging Data

☐ CT Scans (file size approx.. 0.3 GB per subject/series/reconstruction)
- ☐ Inspiratory
- ☐ Expiratory
- ☐ Standard Reconstruction
- ☐ Sharp Reconstruction

☐ Quantitative Data
- ☐ Lung volume, Percent below thresholds, Mean attenuation (Histo)
- ☐ Lobar data, as above
- ☐ Airway measurements- Inspiratory only, unless Expiratory is specified (limited)(
- ☐ LAC

Additional Requests/Comments:

All data to be sent will be de-identified unless a limited dataset request has been approved and data use agreements are signed.

Request for a Limited Data Set (containing subject identifiers/dates)		
	IRB Approved Protocol	Data Use Agreement

Genetic Data		
Imaging Data		
Phenotypic Data		

Method of data transfer: Total size of files will impact possible methods. Total size < 8 GB can be transferred via secure SFTP. *CT scans will require the requestor to supply a portable hard drive.*

Name and Email Address for Data Recipients:

Data Access
Investigator Certification Form
Required prior to transfer of data.

Title of Study: _____
Ancillary Study Number _____ (assigned by COPDGene)
Principal Investigator: _____

COPDGene and _____ (Name of Principal Investigator) hereby enter into this Distribution Agreement.

Investigator: _____, whose principal affiliation is with _____, requests access to Study Data and/or Materials.

Confidentiality and Privacy of Subjects

I certify that all data (genetic, imaging, phenotypic) obtained from COPDGene will be used in a manner that protects the privacy and confidentiality of the participant subjects. No attempt will be made to identify subjects, to link specimens or data to personal information that is readily identifiable, or to contact COPDGene subjects in relation to the specimens, CT images or data.

Restricted Use of Data

I certify that all data obtained from COPDGene will be used for research purposes only, at this institution only, and only for the analyses described in this request. Also, the data will not be allowed to come into the possession of any other persons except those engaged in research under my direct supervision who accept these restrictions.

This Agreement is not transferable. Recipient agrees that substantive changes made to the Research Project described above, and/or appointment by Recipient of another Investigator to complete the Research Project, require execution of a new Agreement in which the new Investigator and/or new Research Project are designated.

Publications and Acknowledgement of COPDGene Support

I agree to follow the COPDGene Publications Policies and Procedures for publications incorporating data arising from use of the COPDGene database and to acknowledge COPDGene in all publications and presentations of studies utilizing data from COPDGene.

Recipient's Resulting Data to be Provided to COPDGene DCC

I agree to provide newly created analytic variables derived from the COPDGene data set to the COPDGene DCC for incorporation into the general data set. Use of new Data generated solely by the Research Project from Data distributed by COPDGene will be restricted for use by the Collaborating Investigator for 12 months after the completion of the Research Project unless other arrangements are mutually agreed upon. After 12 months, the Data generated by the Research Project will be available to all COPDGene Investigators. The COPDGene Investigator(s) agree to acknowledge the contribution of the Collaborating Investigator(s) who generated the Data from the Research Project in any and all oral and written presentations, disclosures, and publications resulting from any and all analyses of Data generated by the Research Project under this agreement. The COPDGene Investigator will acknowledge Collaborating Investigator(s) as co-authors, as appropriate, on any publication.

Reporting Requirements

Recipient agrees to provide an annual report to COPDGene using the Ancillary Study Investigator Annual Progress Report Form.

Protection of Limited Data Sets

I certify that all limited data sets obtained from COPDGene will be properly stored, encrypted and password protected at all times and deleted after the analyses described in this request have been completed.

Requesting Investigator

Signature by the Requesting Investigator is documentation of agreement with Items 1 through 6 above.	
Signature:	Date:
Printed Name:	
Title:	
Telephone:	
E-mail address:	

Appendix II. Biospecimen Request Form

*The costs associated with the transfer of biospecimens are the responsibility of the requestor. Cost estimates can be provided prior to submission of the proposal.

Investigator: _____
Date: _____
Title of Ancillary Study: _____
Ancillary Study Number: _____ (assigned by COPDGene)

Biospecimens – Plasma/Serum							
	GOLD Stage	Number of Patients	Previously thawed acceptable?	Volume requested	Other Inclusion/ Exclusion Criteria	Visit 1	Visit 2
☐ Serum	U ☐						
	0 ☐						
	1 ☐						
	2 ☐						
	3 ☐						
	4 ☐						
	Any☐						
☐ Plasma	U ☐						
	0 ☐						
	1 ☐						
	2 ☐						
	3 ☐						
	4 ☐						
	Any☐						

Biospecimens – DNA						
	GOLD Stage	Number of Patients	Plates or Tubes	Quantity Requested (µg)	Concentration Requested (ng/µl)	Other Inclusion/ Exclusion Criteria
☐ DNA	U ☐					
	0 ☐					
	1 ☐					
	2 ☐					
	3 ☐					

	4 □					
	Any□					

Additional Requests/Comments:

Ship to Address:

_____

_____

_____

Biospecimen Request Investigator Certification

Required prior to shipment of specimens

Title of Study: _____
Principal Investigator: _____
NIH Grant Number (if relevant): _____

COPDGene and _____ (Name of Principal Investigator) hereby enter into this Distribution Agreement.

Investigator: _____, whose principal affiliation is with _____, requests access to Study Data and/or Materials.

Confidentiality and Privacy of Tissue Donor Subjects

I certify that all specimens, images and data obtained from COPDGene will be used in a manner that protects the privacy and confidentiality of the participant subjects. No attempt will be made to identify subjects, to link specimens or data to personal information that is readily identifiable, or to contact COPDGene subjects in relation to the specimens, CT images or data.

Restricted Use of Biological Specimens and Data

I certify that all specimens, images and data obtained from COPDGene, and any materials derived from said specimens, will be used for research purposes only, in my laboratory only, at this institution only, and only for the experiments described in this request (see Appendix II). Also, the specimens or material derived from them and the individual-level data will not be allowed to come into the possession of any other persons except those engaged in research under my direct supervision who accept these restrictions. Recipient agrees that Biological Materials, their progeny, or unmodified or modified derivatives thereof will not be used in human experimentation of any kind.

This Distribution Agreement is not transferable. Recipient agrees that substantive changes made to the Research Project described above, and/or appointment by Recipient of another Investigator to complete the Research Project, require execution of a new Distribution Agreement in which the new Investigator and/or new Research Project are designated.

Publications and Acknowledgement of COPDGene Support

I agree to follow the COPDGene Publications Policies and Procedures for publications incorporating data arising from use of the COPDGene biological specimens requested and to acknowledge COPDGene in all publications and presentations utilizing specimens or data from COPDGene.

Annual Report

I agree to provide an Annual Progress Report regarding my use of the COPDGene biological specimens.

Biosafety

I am aware that all specimens distributed by the Biorepository may be infectious and potentially biohazardous. I understand that the requested specimens may pose health risks to persons handling or in the vicinity of the specimens, the environment, and the community. I certify that I am cognizant of and will employ the appropriate biosafety standards including special practices, equipment, and facilities and will comply with all applicable institution policies and state and federal government health and safety regulations. I will also directly supervise all users of the specimens and will assure that those users are cognizant of and comply with safety standards and good laboratory practices.

Recipient's Resulting Data to be Provided to COPDGene DCC

Recipient agrees to provide the DCC or designee copies of all Data, including Genetic Analysis Data, which are developed based on the biologic specimens distributed by COPDGene, within 12 months of its collection unless other arrangements are mutually agreed upon.

Use of Data or Materials generated solely by the Research Project from biologic specimens distributed by COPDGene will be restricted for use by the Collaborating Investigator for 12 months after the completion of the Research Project unless other arrangements are mutually agreed upon. After 12 months, the Data and Materials generated by the Research Project will be available to all COPDGene Investigators. The COPDGene Investigator(s) agrees to acknowledge the contribution of the Collaborating Investigator(s) who generated the Data and Materials from the Research Project in any and all oral and written presentations, disclosures, and publications resulting from any and all analyses of Data or Materials generated by the Research Project under this agreement. The COPDGene Investigator will acknowledge Collaborating Investigator(s) or designee as co-authors, as appropriate, on any publication.

Costs

Cost for distribution of Biological and other Materials, including DNA, will be determined by the appropriate COPDGene laboratory and should be covered by the investigator.

Return or Destruction of Specimens

I certify that all specimens obtained from COPDGene, and any materials derived from said specimens, will be returned to the CODPGene Biorepository or, if approved by the COPDGene Executive Committee, destroyed after the experiments described in this request have been completed and no additional studies utilizing the same tissues will be performed without review of the new study by the COPDGene Executive Committee.

Requesting Investigator

Signature by the Requesting Investigator is documentation of agreement with Items 1 through 8 above.	
Signature:	Date:
Printed Name:	

Title:
Telephone:
E-mail address:

Investigator Annual Progress Report

Will be incorporated into annual IRB report if biological specimens or a limited dataset are part of this ancillary study.

Investigator: _____

Title of Ancillary Study: _____
Ancillary Study Number _____ (assigned by COPDGene)

Please send this report and direct any questions to:

Elizabeth Regan, M.D., Ph.D., Associate Director, COPDGene, at ReganE@njhealth.org

Date Specimens/Data Received: _____
Date of This Report: _____

PROGRESS ON ANALYSES

	None	In Preparation	Submitted	Accepted*
Presentations	☐	☐	☐	☐
Publications	☐	☐	☐	☐

*If accepted, please send the citation and a paper copy or reprint to the Administrative Core.

Brief Summary of Progress:

Signature:	Date:
Printed Name:	
Title:	
Telephone:	
E-mail address:	

Appendix III. dbGaP Overview and Operations

The data in dbGaP will be pre-competitive, and will not be protected by intellectual property patents. Investigators who agree to the terms of dbGaP data use may not restrict other investigators' use of primary dbGaP data by filing intellectual property patents on it. However, the use of primary data from dbGaP to develop commercial products and tests to meet public health needs is encouraged. Investigators interested in Data use from dbGaP must agree to the dbGaP Certificate of Confidentiality (see attached).

dbGaP Submission Policy

Submissions to dbGaP will not be accepted without assurance that the submitting institution approves the submission and has verified that the data submission is consistent with all applicable laws and regulations, as well as institutional policies. Submitters must also identify any limits on research uses of the data that are specifically set by individual research participants, e.g., through their informed consent. The COPDGene® DSR will be the governing body responsible for ensuring dbGaP submission policies are followed and are consistent with the COPDGene® Study Protocol.

dbGaP Data Content and Organization

1. Open Access Data
Open-access data can be browsed online or downloaded from dbGaP without prior permission or authorization. These data will include, but may not be limited to, the following:

dbGaP Data Type	**Where to Find It**
Studies	'Study' column when browsing studies
	Result of a search under the tab 'Studies'
	Part of the breadcrumb path of a variable or document
Study Documents	Link from 'Browse Studies'
	Link under 'Associated Documents' on study report
	Result of a search under the tab 'Study Documents'
Phenotypic Variables	Link under 'Browse Studies'
	Link under 'Associated Variables' on study report
	Result of a search under the tab 'Variables'
Genotype-Phenotype Analyses	Link under 'Associated Analyses' on variable report
	Link under 'Associated Analyses' on study report

2. Controlled-Access Data
Controlled-access data can only be obtained if a user has been authorized by the appropriate Data Access Committee. Information on requesting controlled data access is available below. Data available to authorized investigators may include the following:

- de-identified phenotypes and genotypes for individual study subjects

- pre-computed univariate associations between genotype and phenotype (if not made available on the public site)

Since data access policies are determined on a per-study basis, data available to users with controlled access authorization may vary between studies and may also change from what is described here without notice.

Requesting Controlled-Access Data through dbGaP

Access to controlled data in dbGaP will be granted by an NIH Data Access Committee (DAC). Users wishing access to controlled data must submit a Data Use Certification (DUC), to the appropriate NIH DAC for approval. DAC approval for controlled data access will be dependent upon completion of the DUC, and confirmation that the proposed research use is consistent with the COPDGene® CDMP oversight and authorization, and the COPDGene® Study protocol.

COPDGene® CDMP submissions of controlled-access data housed in dbGaP may retain the exclusive right to publish an analysis of their submitted data for a specified period of time. Users of controlled-access data will have to consult the DUC to determine the specific publishing exclusivity period for the COPDGene® Study.

Submitting Data to dbGaP

In addition to providing individual-level phenotype and genotype data to dbGaP, dbGaP also requires the submission of sufficient metadata to enable NCBI to provide a browseable interface for a study.

The following will be included in submissions to dbGaP:

- study documents (i.e. manual of procedures, protocols, questionnaires, consent forms, etc.)
- data dictionary (a description of measured variables with pointers to those parts of study documents that describe how variables were measured)
- any other supporting documentation
- phenotype, exposure, and genotype without identifiable information, created using a random, unique code whose key will be held by COPDGene® CDMP.
- a guarantee that the identities of research participants will not be disclosed to dbGaP, or to secondary users of the coded data (due to this guarantee, research participants should not expect the return of individual research results derived from the analyses of submitted data)
- a statement verifying that the data submitted to dbGaP for subsequent sharing and appropriate research is consistent with the initial informed consent process completed by study participants of COPDGene®.
- a statement identifying any uses of the data that are specifically excluded by the informed consent process
- a statement from the COPDGene® DSR that submission of the data is in accord with all applicable laws and regulations

As dbGaP is a NCBI data distribution service, the control and management of the data housed in dbGaP is under the jurisdiction of the COPDGene® DSR; therefore, any questions regarding submission requirements or other data issues should be directed to the DAC for the study in question.

Definitions

Data Access Committee (DAC): Data Access Committees are established based on programmatic areas of interest as well as technical and ethical expertise. All DACs will operate through common principles and under similar mechanisms to ensure the consistency and transparency of the controlled- data access process.

Data Use Certification (DUC): A Data Use Certification is the application a user submits to a particular study's Data Access Committee (DAC) for consideration for authorized use of controlled dbGaP data. The Data Use Certification should include a list of the controlled data set(s) required by the user and a brief description of the proposed research use of the requested data. The user must also offer the following assurances in the Data Use Certification that:

- the data will only be used for approved research
- data confidentiality will be protected
- all applicable laws, local institutional policies, and terms and procedures specific to the study's data access policy for handling dbGaP data will be followed
- no attempts will be made to identify individual study participants from whom data were obtained
- controlled-access data from dbGaP will not be sold or shared with third parties
- the contributing investigator(s) who conducted the original study and the funding organizations involved in supporting the original study will be acknowledged in publications resulting from the analysis of those data
- all NIH supported genotype/phenotype data and conclusions derived directly from them will remain in the public domain, without licensing requirements
- an annual research progress report will be submitted to the study's DAC

Appendix IV. COPDGene Cores and Clinical Centers

COPDGene Study Cores

Administrative Core – *National Jewish Health*

Data Coordinating Center – *National Jewish Health*

Genetic Analysis Core – *Brigham and Women's Hospital, Johns Hopkins University, Harvard School of Public Health, University of Colorado Denver, National Jewish Health*

Sequencing and Bioinformatics Core – *Brigham and Women's Hospital, Harvard School of Public Health*

Sample Storage Core (Biorepository) – *Brigham and Women's Hospital, Johns Hopkins University*

Imaging Core – *National Jewish Health, Brigham and Women's Hospital, University of Iowa*

Epidemiology Core – *University of Colorado Denver*

Subtyping Core – *Brigham and Women's Hospital, Northeastern University, Harvard School of Public Health, University of Colorado Denver*

Biomarkers Core – *National Jewish Health, Brigham and Women's Hospital*

Mortality Adjudication Core – *National Jewish Health*

COPDGene Study Clinical Centers

Ann Arbor VA Medical Center *(AVA)*

Jeffrey L. Curtis, MD
Pulmonary & Critical Care Medicine
University of Michigan Health System
Ann Arbor, MI 48109
Tel: 734-845-3457
Fax: 734-845-3257
Beeper: 734-936-6266 #3082
jlcurtis@umich.edu

Baylor College of Medicine *(BAY)*

Nicola A. Hanania, MD, MS
Baylor College of Medicine
6620 Main St. Suite 11A.01.4
Houston, TX 77030
Tel: 713 873 3454; 713 798 2347
Fax: 713 873 3346
hanania@bcm.tmc.edu

Brigham and Women's Hospital *(BWH)*

Dawn L. DeMeo, MD, MPH
Brigham and Women's Hospital
Channing Laboratory
181 Longwood Avenue
Boston, MA 02115
Tel: 617-525-0866
Pager 617-732-6660, #33904
redld@channing.harvard.edu

Craig P. Hersh, MD, MPH
Channing Laboratory
Brigham and Women's Hospital
181 Longwood Ave.
Boston, MA 02115
Tel: 617-525-0729
Fax: 617-525-0958
craig.hersh@channing.harvard.edu

Columbia University Medical Center *(COL)*

R. Graham Barr, MD, DrPH
Departments of Medicine and Epidemiology
Columbia University Medical Center
PH 9 East Room 105
630 West 168th St, New York, NY 10032
Tel: 212-305-4895
Fax: 212-305-9349
rgb9@columbia.edu

Duke University Medical Center *(DUK)*

Neil MacIntyre, Jr., MD
Erwin Road, Room 7453
Box 3911
Duke University Medical Center
Durham, NC 27710
Tel: 919-681-5691
Fax: 919-681-2892
neil.macintyre@duke.edu

Johns Hopkins University *(JHU)*

Robert A. Wise, MD
Johns Hopkins Bloomberg School of Public Health
Asthma Ctr 4B72
615 North Wolfe Street
Baltimore, Maryland 21205
Tel: 410-550-0506 or 410-550-0545
Fax: 410-550-2612
rwise@jhmi.edu

Nadia Hansel, MD, MPH
Division of Pulmonary and Critical Care
Johns Hopkins University
1830 E. Monument St., 5th floor
Baltimore, MD 21205
Tel: 410-502-7041; 410-550-2935
Pager: 410-283-6552
nhansel1@jhmi.edu

L.A. Biomedical Research Institute *(HAR)*

Richard Casaburi, MD
Associate Chief, Division of Respiratory
Los Angeles Biomedical Research Institute
at Harbor-UCLA Medical Center
1124 W. Carson St., Bldg J4
Torrance CA 90509
Tel: 310-222-8200
Fax: 310-328-8249
casaburi@ucla.edu

Michael E. DeBakey VAMC *(HVA)*

Amir Sharafkhaneh, MD
Michael E. DeBakey VAMC
2002 Holcombe Blvd., Rm. 3C-220
Houston, TX. 77030
Tel: 713-794-7318
Fax: 713-794-7295
amirs@bcm.tmc.edu

Minneapolis VA Medical Center *(MVA)*

Chris H. Wendt, MD
University of Minnesota
PACC Medicine
350 VCRC
401 E River Rd
Minneapolis, MN 55455
Tel: 612-624-5682
Fax: 612-625-2174
wendt005@umn.edu

Minnesota Health Partners *(HPR)*

Charlene E. McEvoy, MD, MPH
HealthPartners Research Foundation
401 Phalen Blvd
St. Paul, MN 55130
Tel: 952-967-5493
Fax: 651-254-7676
charlene.e.mcevoy@healthpartners.com

Morehouse School of Medicine *(MSM)*

Marilyn G. Foreman, MD, MS
Morehouse School of Medicine
720 Westview Dr. SW
Atlanta, GA 30310
Tel: 404-616-4658
Fax: 404-616-5474
Pgr: 404-415-0145
mforeman@msm.edu

National Jewish Health *(NJC)*

Russell P. Bowler, MD, PhD
National Jewish Health
1400 Jackson Street, K715a
Denver, CO 80206
Tel: 303-398-1639
Fax: 303-270-2249
Bowlerr@njhealth.org

Reliant Medical Group *(FAL)*

Richard Rosiello, MD
Reliant Medical Group
123 Summer St.
Worcester, MA 01608
Tel: 508-368-3120
Fax: 508-368-3121
richard.rosiello@reliantmedicalgroup.org

Temple University *(TEM)*

Gerard J. Criner, MD
Temple University
3401 N Broad St, 7th Fl
Philadelphia, PA 19140-5103
Tel: 215-707-8113
Fax: 215-707-6867
gerard.criner@temple.edu;
CrinerG@tuhs.temple.edu

Univ. of Alabama, Birmingham *(UAB)*

Mark T. Dransfield, MD
Division of Pulmonary, Allergy and Critical Care Medicine
University of Alabama at Birmingham & Birmingham VA Medical Center
1900 University Blvd, 215 THT
Birmingham, AL 35294
Tel: 205-934-7557
mdransfield99@msn.com

Univ. of California, San Diego *(USD)*

Joe W. Ramsdell, MD
University of California, San Diego
200 W. Arbor Drive # 8415
San Diego, CA 92103-8415
Tel: 619-543-6275 (select 3)
Alternate: 619-543-7241
Fax: 619-543-3383
jramsdell@ucsd.edu

University of Iowa *(UIA)*

Alejandro Comellas, MD
University of Iowa Hospitals and Clinics
200 Hawkins Drive
Internal Medicine/Pulmonary
C 331 GH
Iowa City, Iowa 52242
Tel: 319-384-6484
alejandro-comellas@uiowa.edu

University of Michigan *(UMC)*

MeiLan K. Han, MD, MS
3916 Taubman Center
1500 E Medical Center Drive
University of Michigan Health System
Ann Arbor, MI 48109
Tel: 734-615-9772
Cell: 734-645-3120
mrking@umich.edu

Univ. of Minnesota *(UMN)*

Joanne Billings, MD, MPH
COPDGene® Clinical Center Director
Pulmonary, Allergy, Critical Care and Sleep Medicine
University of Minnesota
Mayo Mail Code 276
420 Delaware Street S.E.
Minneapolis, Mn 55455
Tel: 612-624-0999
Fax: 612-625-2174
Billi001@umn.edu

Univ. of Pittsburgh *(PIT)*

Frank Sciurba, MD
Kaufmann Bldg.
University of Pittsburgh Medical Center
3471 Fifth Ave., Suite 1211
Pittsburgh, PA 15213-3227
Tel: 412-648-6494
Fax: 412-692-4842
sciurbafc@upmc.edu

Univ. of Texas Health Science Center at San Antonio *(TEX)*

Antonio Anzueto, MD
UTHSC at San Antonio
South Texas Veterans Health Care System
7400 Merton Minter Drive, Room C516.1
San Antonio, TX 78229-4404
Tel: 210-617-5256
Fax: 210-949-3006
anzueto@uthscsa.edu

Appendix V. COPDGene Ancillary Study Proposal

Ancillary Studies Policies and Procedures

Ancillary Studies are managed by the COPDGene Ancillary Studies and Publications Committee.

COPDGene Ancillary Studies and Publications Committee

James Crapo, MD, Co-Chair
John Hokanson, PhD, Co-Chair
Jeffrey Curtis, MD
Dawn DeMeo, MD
Mark Dransfield, MD
Doug Everett, PhD
Marilyn Foreman, MD
MeiLan Han, MD, MS
Craig Hersh, MD
Victor Kim, MD
David Lynch, MD
Barry Make, MD
Elizabeth Regan, MD, PhD
Edwin Silverman, MD, PhD
Matt Strand, PhD
Executive Secretary: Sara Penchev
Staff: Carla Wilson

Mission:
The mission of the COPDGene Ancillary Studies and Publications Committee is to facilitate research using the COPDGene cohort and to facilitate the rapid publication and presentation of the highest quality research from the COPDGene investigator teams.

This Policies and procedures document will include the following appendices:

- COPDGene Ancillary Study Proposal Requirements
- Data Sharing and Access Policy, Data Request Form, and Data Access Investigator Certification Form
- Biospecimen Request Form & Biospecimen Request Investigator Certification
- Ancillary Studies Investigator Annual Progress Report

Definition of an Ancillary Study:
A COPDGene ancillary study is one that extends research questions or resources beyond that in the original program. Several types of ancillary studies will be included:

1) Analytical Ancillary Studies: Additional analyses of the main study data which are not included in the main analyses of the study;

2) Supplemental Ancillary Studies: Use of COPDGene study data from one or more clinical centers in concert with additional data collection at those clinical centers on their study participants;

3) Related Ancillary Studies: Ongoing studies which include COPDGene study participants at Clinical Centers that may overlap with main study goals (e.g., candidate gene studies, radiology studies).

All three of these types of ancillary studies will require review and approval by the Ancillary Studies and Publications Committee. It may be done as a component of the primary COPDGene funding or may derive funding from other than COPDGene funds. Examples include studies funded by investigator-initiated NIH research awards (R01s), grants from academic institutions, private sources (e.g., drug companies), or those performed at no cost (generally because of the special interest of a researcher). Ancillary studies involve the collection of new data, either directly from participants or from previously collected samples, images, or other sources (e.g., medical records).

Philosophy:
Investigators individually and collaboratively are encouraged to propose and conduct ancillary studies. Such studies enhance the value of COPDGene and ensure the continued interest of the diverse group of investigators who are critical to the successes of the study as a whole. They provide an exceptional opportunity for investigators, both within and outside of COPDGene, to conduct additional projects at minimal cost.

At each level of review, highest priority will be given to studies that:

1. Do not interfere with the main COPDGene objectives
2. Have the highest scientific merit
3. Produce the smallest burden on COPDGene participants and the least demand on COPDGene resources, such as blood samples
4. Demonstrate willingness to follow the COPDGene guideline for data sharing
5. Have the potential to develop new sources of funding to exploit the opportunities created by the COPDGene core program
6. Require the unique characteristics of the COPDGene cohort
7. Demonstrate a feasible plan for funding additional costs of the study.

Necessary Approvals:
The COPDGene Ancillary Studies and Publications Committee and Executive Committee must approve ancillary study proposals prior to implementation. The COPDGene Ancillary Studies and Publications Committee provides initial review and makes recommendations to the Executive Committee. Once an ancillary study is approved, the investigator will receive an approval letter from the Executive Secretary of the Ancillary Studies and Publications Committee. All approved COPDGene Ancillary Studies are tracked and a list is available on the COPDGene website: https://dccweb.njhealth.org/sec/COPDGene/rptAncillaryStudies.cfm.

Responsibilities of Ancillary Study Investigators:

1. Costs: The investigator applying for an ancillary study may be required to provide the additional funds required to conduct the study. The Executive Committee will determine whether or not proposed ancillary studies can be supported in whole or in part by the COPDGene primary study resources. The Ancillary Studies and Publications Committee and the Executive Committee will be concerned with both the

obvious and the hidden costs to COPDGene entailed by an ancillary study (such as costs to the Data Coordinating Center for coordinating the additional data collection, costs to Clinical Centers, and costs to our Biological Repository for retrieving samples, etc). The costs associated with the transfer of data and/or Biospecimens are the responsibility of the requestor. Cost estimates can be provided prior to submission of the proposal.

2. Additional Funding: PIs who plan to propose an ancillary study with the intention of seeking grant funding should consult with the COPDGene Administrative Core to determine what level of involvement will be required of the COPDGene Program and the associated costs. In general, this will result in a subcontract proposal to be included in the PI's grant application. COPDGene will provide a letter of support when requested, for investigators applying for funding.
3. Confidentiality and Identification of COPDGene Participants: Confidentiality of individually identifiable data about COPDGene participants must be assured. As a general rule, no personal identification of participants will be provided to ancillary studies staff. There are no assurances that participants will be able to be identified and contacted in the future for the purposes of an ancillary study.
4. Clinical Implications of Findings: The proposing investigator must clearly delineate any findings of clinical significance that may result from the study, including genetic findings, and propose how these will be handled, including reporting to participants and their physicians and providing recommendations for follow up. This includes incidental findings, such as pathology identified from an imaging study that is not the focus of the study.
5. Genetic Studies: Ancillary studies should complement and not overlap with the primary research goals of the COPDGene study.
6. Inclusion of COPDGene Investigator(s): A COPDGene investigator must be included as a co-investigator on an ancillary study. This individual is responsible for presenting the study to the Ancillary Studies and Publications Committee, monitoring the study to assure continuing compatibility with COPDGene and serving as a liaison to the COPDGene Ancillary Studies and Publications Committee. In addition, each manuscript and abstract is expected to include a COPDGene investigator.
7. IRB Approval: The appropriate institution review boards must approve all ancillary studies before they are performed. IRB approval is not required to submit a proposal for an ancillary study to COPDGene.
8. Final Application or Proposal: For ancillary studies involving a new grant application, a copy of the final proposal as submitted for funding should be sent to the COPDGene Administrative Core.
9. Industry Participation: Proposals for industry sponsorship or collaboration are encouraged. It will be the responsibility of the PI of the ancillary study to obtain agreement through an appropriate contractual mechanism that all data relevant to the COPDGene ancillary study will follow the open access and data sharing policies of COPDGene.

10. Status Reports: The ancillary study PI is responsible to keep the COPDGene Administrative Core apprised of major developments in the life of the application or proposal, including success of funding, start date, changes in protocol, completion and any resulting publications or presentations.
11. Review of Publications and Presentations: Manuscript proposals based on approved ancillary studies should be submitted to and approved by the COPDGene Ancillary Studies and Publications Committee. All publications, presentations and abstracts from an approved ancillary study should be reviewed and approved by the COPDGene Ancillary Studies and Publications Committee prior to submission or presentation, in accordance with the COPDGene general rules for publications and presentations and the COPDGene Publications Policies and Procedures document.
12. Termination of Ancillary Study: The Executive Committee by majority vote may terminate an ancillary study if it judges that the study has become too burdensome, its scientific value has diminished, or it has failed to make substantial progress toward completion of its goals.

COPDGene Ancillary Study Review Procedures

1. Principal Investigator should submit the ancillary study proposal to the COPDGene Administrative Core Executive Secretary(c/o Sara Penchev, PenchevS@NJHealth.org). The Ancillary Studies Policies and Procedures will be posted on the open access portion of our study web site.

   https://dccweb.njhealth.org/sec/COPDGene/Ancillary.cfm

   Questions should be directed to the COPDGene Administrative Core (c/o Sara Penchev).

   Ancillary Study meetings will be scheduled the first and third Thursday of every month. Complete proposal forms (including Data and Biospecimens requests) should be sent to the Executive Secretary no later than the Monday before each meeting.

   COPDGene website access will be needed to access the policies and access form. If one does not have website access, the COPDGene sponsor of the proposal may submit the form on the requester's behalf.

   Prior to submitting a proposal, investigators should access the below website to search title key words to identify potential for overlap:

   https://dccweb.njhealth.org/sec/COPDGene/ListAncillaryStudies.cfm

   To search the list, select Control +F on your keyboard and search key words individually.

2. The COPDGene Administrative Core will conduct an initial review of proposals for administrative compliance and to determine involvement of other COPDGene Centers. If the proposal is not complete, it will be returned by e-mail to the investigator for revision and resubmission.
3. The COPDGene Administrative Core will forward the proposal by e-mail to the COPDGene Ancillary Studies and Publications Committee and to relevant COPDGene centers and/or committees with the meeting agenda on the Monday prior to the meeting. . The chairs of the Ancillary Studies and Publications Committee will review the proposals on a conference call or will handle the review by e-mail.
4. Ancillary Study proposals approved by a majority vote of the Ancillary Studies and Publications Committee will be discussed by the Executive Committee during its regular weekly conference calls. The Ancillary Studies and Publications Committee and/or the Executive Committee may also invite the PI (and/or the PI's COPDGene sponsor) to present the proposal and answer questions and then absent him/herself during the subsequent discussion and voting. Once the Ancillary Study is approved, the Administrative Core Executive Secretary will send an approval letter to the requester.
5. If the proposal requires revisions, the comments of the Ancillary Studies and Publications Committee (and Executive Committee, if applicable) will be sent to the PI. The PI must address these comments in a separate letter that accompanies the revised proposal and send these to the COPDGene Administrative Core Executive Secretary. Revised proposals will be reviewed by the Ancillary Studies and Publications Committee at the next meeting. If approved by the Ancillary Studies and Publications Committee, the revised proposal will go to the Executive Committee where a majority vote will be the basis for a final decision. A proposed ancillary study which is returned for revision will be considered to be withdrawn if a revised proposal is not received within 12 months.
6. Submission and receipt date of all Ancillary Study Proposals will be tracked by the COPDGene® Administrative Core. The Administrative Core will also maintain the submission proposal form, approval letter, protocol, and other supporting documents as applicable in a restricted access file depot. A list of all ancillary studies approved by the COPDGene Ancillary Studies and Publications Committee will be maintained on the COPDGene website.
7. **Approval of an Ancillary Study does not automatically include approval of resulting publications. Manuscript proposals for results arising from this ancillary study should be submitted to the COPDGene Ancillary Studies and Publications Committee in accordance with the COPDGene Publications Policies and Procedures.**

COPDGene Ancillary Study Proposal Requirements

Submit proposal to the COPDGene Administrative Core (c/o Sara Penchev, PenchevS@NJHealth.org). For scientific and technical questions regarding this application, contact Dr. John Hokanson (john.hokanson@ucdenver.edu).

PART I: Basic Information and Checklist of Key Components

1. Date of Submission:
2. Proposed Title:
3. Proposing Investigator's Name and Contact Information (include email):
4. COPDGene Sponsor (if different from above):
5. Co-author Names and Email Addresses (Note: All proposals will be made available to COPDGene investigators to allow for additional interested co-authors):

6. Have all co-authors reviewed and approved this document? (required):
   ☐ Yes ☐ No

PART II: Description (Please limit this section to 2 pages if possible.)

1. Research Question with Hypothesis and Specific Aims:
2. Research Design – Primary Methods and Procedures to be Employed:
3. Brief Analysis Plan and Methods:
4. Impact on the Main Study. For example:
   a. Burden on study participants
   b. Impact on DCC
   c. Impact on Core resources
5. Number and Name of Clinical Sites to be Involved and Number of Subjects Needed:
6. Projected Costs:
   a. Data Coordinating Center
   b. Imaging Core
   c. Genetics Core
   d. Biorepository
   e. Other
7. Are any Biological Specimens requested? If so, the separate COPDGene Biospecimen Request Form must be completed and also indicate when in the protocol this will be done. Appendix C of this document.

8. Is genetic, imaging or phenotypic data requested? If so, the separate COPDGene Data Request Form must be completed.
9. How will the Ancillary Study be funded?
10. Will data from this study be made publicly available?
11. Will the study use the main COPDGene consent, an amended COPDGene consent, or a separate consent? Attach copy of proposed consent.
12. List and explain all additional data to be collected.

Appendix VI. Crowd Sourcing Ancillary Study

We would like to propose the following crowdsourcing approach to data analysis and innovation in the "Genetic Epidemiology of COPDGene" (COPDGene) project which is included in National Jewish IRB Protocol Number 2278. This is a collaborative study between the COPDGene investigators and Pfizer.

Purpose

Open innovation and contest-based crowdsourcing challenges using clinical, genetics and/or phenotypic data already collected from the COPDGene cohort will be used to develop algorithms to solve tough analytic questions. By accessing the collective knowledge of the "crowd" we hope to develop innovative and novel solutions to analytical questions associated with "big data" including patient level medical data. Our goal is to perform these analyses without compromising the identifiability of study participants in any way.

In general we aim to harness open innovation through crowdsourcing to probe one or more of the following general categories of questions:

a) Provide an unbiased assessment of models and methodologies for the prediction of COPD phenotypes using clinical, phenotype and/or genetics data.
b) Identify a diagnostic signature for Chronic Obstructive Pulmonary Disease (COPD) phenotypes (classification) using clinical, phenotype and/or genetics data.
c) Automate image feature extraction for 3D CT scans.
d) Apply supervised learning via machine learning algorithms to explore the correlation of clinical, phenotype and/or genetics data to identify novel patterns in the data leading to greater insights into disease understanding.

Background and significance

This application is being applied as an addendum to the COPDGene study. COPDGene has been established to understand the causes of COPD and COPD heterogeneity, to reclassify COPD based on the etiology of the disease, and to evaluate longitudinal progression in the COPD subjects. The goal for this addendum is explore additional data analytical methods to help understand and better interpret the features of the disease that lead to disease progression and to identify algorithms that predict patients that are more susceptible to disease progression.

Advances in biotechnology have fueled the generation of unprecedented quantities of data across the life sciences. However, finding analysts who can address such 'big data' problems effectively has become a significant research bottleneck. Historically, prize-based contests have had striking success in attracting unconventional individuals who can overcome difficult challenges. In academia and elsewhere, this bottleneck is more than just a personnel shortage. Available personnel may lack experience with the specific approaches or techniques required.

As an alternative to an extensive search to identify and contract with potentially suitable analysts, prize-based contests have emerged as a novel approach to find solutions to challenging

problems in settings as diverse as industrial R&D, software development and internet commerce. Such contests are one part of a decade-long trend toward solving science problems through large-scale mobilization of individuals by what the popular press refers to as 'crowdsourcing'. In general, crowdsourcing has come to imply a strategy that relies on external, unaffiliated actors to resolve a particular problem.

Over the last ten years, online prize-based contest platforms have emerged to solve specific scientific and computational problems for the commercial sector. These platforms, with solvers in the range of tens to hundreds of thousands, have achieved considerable success by exposing thousands of problems to larger numbers of heterogeneous problem-solvers and by appealing to a wide range of motivations to exert effort and create innovative solutions. The large number of entrants in prize-based contests increases the probability that an 'extreme-value' (or maximally performing) solution can be found through multiple independent trials; this is also known as a parallel-search process. In contrast to traditional approaches, in which experts are predefined and preselected, contest participants self-select to address problems and typically have diverse knowledge, skills and experience that would be virtually impossible to duplicate locally. Thus, the contest sponsor can identify an appropriate solution by allowing many individuals to participate and observing the best performance. This is particularly useful for highly uncertain innovation problems in which prediction of the best solver or approach may be difficult and the best person to solve one problem may be unsuitable for another.

Design and Methodology

The COPDGene Study has recruited 10,300 subjects stratified by severity of COPD and smoking status to conduct cross-sectional case-control studies. COPDGene has identified and characterized COPD cases and control subjects from two racial groups (non-Hispanic whites and non-Hispanic African Americans) for genetic, epidemiologic, and natural history studies. The COPDGene cohort is comprised of current or former smokers, enrolled between the ages of 45 and 80 years, with a minimum ten pack-year smoking history, as well as a group of similar aged non-smokers.

Data from Phase 1 has been collected and stored in the Data Coordinating Center as de-identified data and a data-sharing plan has been implemented. De-identified data has been made available through dbGaP http://www.ncbi.nlm.nih.gov/gap and to study investigators for analysis.

Using this already collected and de-identified data, we will perform additional alterations to the data to limit any identifiability of the COPDGene subjects. This updated data set will be used to run a series of prize-based analytic challenges on the TopCoder platform, CrowdAnalytix platform, or comparable provider. TopCoder and CrownAnalytix are companies which administers contests in computer programming and contain a community of extremely-talented analysts who regularly compete to solve coding and 'big data' problems. Effective contests require that contestants have access to patient level data. The original data obtained by COPDGene has already been de-identified and stripped of any personal information by the COPDGene DCC and as a result this secondary study will have not have any access to personal information and no personal information will be revealed during this process. However, there are concerns that research subjects could be identified by their genetic variants or their specific

clinical test values (e.g. FEV1 level) alone. Additionally, the original patient consent form did not include public release of the data. In order to mitigate privacy risks to individuals, we will perform a second de-identification prior to the contest, so that contestants will not be able to identify the source for the data. In addition, to further mitigate privacy risks, all data will transformed to mask the type of data collected and the values of the data collected. Transformation algorithms have been created and validated using ciphering experts (accessible through the TopCoder or other platform). Final transformation on the COPDGene data will occur behind the Pfizer firewall and will include the following transformations:

- All data labels will be stripped.
- All continuous data values will be normalized to values between 0.0 and 1.0
- All non-continuous data (categorical data) will also be transformed to integer values
- Where possible columns will be randomized and rearranged
- De-identified patient values such as subject IDs, sites etc. (consistent with HIPAA) will be subjected to a second de-identification protocol or will be removed entirely so that values will never be exposed
- Any genetic biomarkers required will be stripped of labels and used with the values 0, 1 or 2
- The sponsors of the contest including COPDGene and Pfizer will not be mentioned to any of the contestants.
- Contest questions will be reframed and translated so contestants will not be informed of the specific questions being addressed.

Additionally, any genetic information used would be limited to less than 1000 of the ~700,000 markers (SNPs) collected and would include one or more of the following: known lung function markers (75 loci), known COPD susceptibility loci (6-10 loci), known function SNPs (lung eQTL), known SNPs associated with interesting phenotypes (autoimmune diseases), SNPs associated with genes thought to be involved in disease progression, SNPs associated with exacerbation and/or other potential interesting COPD phenotypes.
A portion of the transformed data will be stored on the appropriate vendors (TopCoder.com, crowdanalytix.com, or comparable vendor) site for the duration of the competition, expected to last no longer then 4 weeks. Upon completion of the contest, data will be removed and no longer available. All contestants will be required to sign up with the appropriate vendor (TopCoder.com, crowdanalytix.com, or comparable vendor) site and sign an end user license document for each individual contest entered.

Implicit in the design of the study is that all data and information will be de-contextualized and the question will be presented to the crowd as a purely mathematical problem. Therefore no understanding of medical research would be required to solve these problems. By de-contextualizing the data, we hope to attract potential solvers from different disciplines including but not limited to mathematicians, computer engineers, economists, and physicists, who may have similar experiences in their respective disciplines to apply to our problem. Additionally, through crowdsourcing, we have the potential to access the collective expertise of individual experts in machine learning techniques and algorithms. There are well over a hundred different machine learning algorithms, and it is not clear which algorithm would be best for this dataset. It is impractical to find an expert in each method. Through crowdsourcing a diverse set of experts will explore the solution landscape and test their methods to determine the best algorithm for our data set.

Although the data for this secondary study have been de-identified and transformed, it will be important for scientific reasons that the investigators in this study at COPDGene and Pfizer are able to recreate our transformed data. A key code will be generated to allow re-association and re-transformation of the data. The codes will be stored in a secure environment that protects from loss, theft and any unauthorized access, disclosure, copying, or use. In addition re transformed data will only be available to the principal investigators of this study.

Analysis will be conducted through one of the COPDGene Industry partners, Pfizer. Data protection and privacy concerns will be consistent with Pfizer policies and privacy laws consistent with Pfizer policy REG09-POL (Protecting the Privacy of Personal Information). According to Pfizer policy, personal information may be used without specific notice or consent in a de-identified data format, provided that applicable legal requirements for de-identifying personal information are satisfied by institutional review board (IRB), independent ethics committee (IEC) or data protection authority permission.

In the absence of individual consent through the informed consent process, we are seeking to obtain the right to process patient level data through an alternative mechanism, which includes de-identification and permission by IRB.

Our secondary study would serve a significant public health interest, as any methods we develop would have broad utility to the greater COPD population. Additionally, our secondary study proposes to answer questions consistent with the original use of the data. We have placed additional and rigorous safeguards in place to mitigate patient privacy risks to individuals and to protect the data from inadvertent or unintentional use or disclosure. Unreasonable effort would be required to locate individuals in order to obtain consent.

Planned Data Analysis

Only de-identified and transformed data will be available for analysis. The contestants will train on a portion of the data, and their algorithms will be evaluated, via a scoring metric, on a different portion of the data that the algorithms have not been trained on. For each analytic competition approximately 50% ± 15% of the applicable data will be made available to the contestants for training, the remaining data will be withheld for testing and objective scoring. For prediction-based contests an objective scoring method based on average precision will be used. Final algorithms will be tested behind the Pfizer firewall on the entire untransformed set of data.