PVDOMICS
A Multi-Center Study to Improve Understanding of Pulmonary Vascular Disease Through Phenomics


The National Institutes of Health (NIH)/National Heart, Lung and Blood institute (NHLBI) launched an initiative, PVDOMICS (Redefining Pulmonary Hypertension through Pulmonary Vascular Disease Phenomics) that aims to augment the current pulmonary hypertension (PH) classification based on shared biological features. PVDOMICS will enroll 1500 participants with PH and disease and healthy comparators. Enrollees will undergo deep clinical phenotyping, and blood will be acquired for comprehensive omic analyses that will focus on discovery of molecular-based subtypes of pulmonary vascular disease (PVD) through application of high dimensional model-based clustering methods. In addition to an updated, molecular classification of PVD, the phenomic data generated will be a rich resource to the broad community of heart and lung disease investigators.

Editorial, see p 1106

PH is a hemodynamic condition that causes increased blood pressure in the pulmonary arteries and the right heart leading to adverse clinical outcomes. The current World Symposium on Pulmonary Hypertension (WSPH) classification of PH is based on a combination of patient characteristics, clinical features, and cardiopulmonary hemodynamics, and these features are used to inform treatment options. Aside from heritable pulmonary arterial hypertension, this classification is not tied to molecular or cellular pathobiologic mechanism to explain the pathogenesis of PH. The NIH has a vested interest in understanding the causes and natural history of PH, as well as the discovery of effective treatment options. Since the first large NIH registry of patients with pulmonary arterial hypertension >30 years ago, significant advances in scientific knowledge and translational medicine have occurred, highlighting a need for updating the current clinical classification system.

The NHLBI has sponsored several workshops focusing on PVD research strategic planning over the past decade. PVD encompasses PH and PVD without PH, for example, pulmonary vasculitis and pathological pulmonary vascular remodeling without hemodynamic criteria for PH. Experts identified the need for aggressive genotype–phenotype characterization of PVD and incorporation of systems biology approaches to identify previously unrecognized relationships at a molecular, cellular, or clinical level and to identify potential subjects for appropriate clinical studies. A workshop convened by the American Thoracic Society (ATS) had similar recommendations. Both the NIH and ATS workshops highlighted the recent rapid advances in the mechanistic understanding of PVD, improved imaging techniques, omics (hereafter referred to as -omic) technologies, and novel methods for biomarker discovery that present a unique opportunity for clinicians and scientists working jointly to augment PH classification on the basis of these new metrics.

In response to the identified opportunities for improved PVD research and treatment, a new PVD initiative from the Division of Lung Diseases was launched in 2014 by the NHLBI. Called PVDOMICS, this program (RFA-HL-14–027 and RFA-HL-14–030) supports a national team of multidisciplinary investigators to perform comprehensive, deep phenotyping across groups 1 to 5 PH. A Data Coordinating Center (Cleveland Clinic) and 6 clinical centers (Brigham and Women’s Hospital, Columbia University/Cornell University, Johns Hopkins University, Mayo Clinic, University of Arizona, and Vanderbilt University) were funded with support of the NHLBI and the Pulmonary Hypertension Association.
**Study Hypotheses**

The general hypothesis addressed by the PVDOMICS Study protocol is that epidemiological, biological, and hemodynamic features will allow differentiation of phenotypic similarities and differences among current WSPH group categories (Table). We hypothesize that the integration of clinical metrics with detailed -omic measures in the blood will allow new understanding of PVD and potentially facilitate an updated classification scheme for PVD. In addition to the common hypotheses addressed by the protocol, each center has a specific hypothesis (Online Table I). These studies aim to lead to a newer classification scheme of PH based on shared biological features. PVDOMICS will measure and integrate genomics, transcriptomics, proteomics, metabolomics, coagulomics, cell biomics and tissue functioning, imaging, and other clinical traits deemed necessary to fully describe a PVD phenotype across all PH group categories. Furthermore, through enrollment and detailed clinical and molecular phenotyping in healthy and disease comparator groups across the United States, knowledge of clinical and molecular risk factors, PVD markers, and potential disease mediators will be identified, and the breadth of disease will be understood through collaboration across multiple centers. Overall, PVDOMICS aims to update the classification of PVD using -omic techniques and foster novel biological concepts to catalyze associated innovative research into preventing PVD and developing more efficacious, precision approaches to individual therapy. The study is registered at www.clinicaltrials.gov as NCT02980887.

**Study Protocol**

Please see the Online Data Supplement for the organizational structure, inclusion and exclusion criteria, adjudication plan, and timeline (Online Figure I).

**Table. Specific Hypotheses**

<table>
<thead>
<tr>
<th></th>
<th>Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The molecular basis of pulmonary vascular disease of all pathogeneses will be discovered by integration of biological markers with careful phenotyping of all patients with PH and comparing this data with healthy subjects and with non-PH patients as diseased comparators, such as emphysema and interstitial lung disease.</td>
</tr>
<tr>
<td>2</td>
<td>Racial and sex-related (or ancestry) genetic variation in phenotype, natural history, and responses to therapy will be discovered and lead to more precise diagnostic and therapeutic approaches.</td>
</tr>
<tr>
<td>3</td>
<td>The response, adaptation, and dysfunction of the RV will be elucidated by careful phenotyping, including specialized imaging and -omic correlations.</td>
</tr>
<tr>
<td>4</td>
<td>Exercise pathophysiology will lead to improved early diagnosis of PVD, elucidation of RV-pulmonary vascular interactions, RV functional reserve, failure, and response to therapies.</td>
</tr>
<tr>
<td>5</td>
<td>The biological and genetic features of patients with combined pulmonary venous hypertension and pulmonary arterial hypertension will lead to better differentiation of these 2 pathogeneses of PH, and of shared biological mechanisms.</td>
</tr>
<tr>
<td>6</td>
<td>Epigenetic and RNA variants will influence the development, severity, and type of PVD and reveal therapeutic responses and form a basis for new therapies.</td>
</tr>
</tbody>
</table>

PH indicates pulmonary hypertension; PVD, pulmonary vascular disease; and RV, right ventricle.

A total of 1500 PH participants, PVD at risk comparators, and healthy controls will be recruited from the clinical centers. The study will recruit participants who present for evaluation of PH, heart failure, lung disease, dyspnea, and exercise intolerance. After the right heart catheterization, participants with PH will be classified according to the traditional WSPH PH groups 1 to 5 for adjudication purposes. Alternatively, patients with mild elevation in mean pulmonary arterial pressure (20–25 mmHg) or normal pulmonary arterial pressure will be assigned to WSPH 1 to 4 comparator groups who are at risk for PVD with similar underlying diseases or as healthy controls. Target recruitment goals for both patients with PH and comparator groups across all centers are in Online Tables II and III.

Participants may be referred to the study centers for enrollment and will undergo a clinical phenotyping protocol that includes a rigorous assessment of pulmonary and right ventricular structure and function with collection of biospecimens for -omic analysis (Figure). Subjects will undergo quality of life assessment, sleep study, body compositional analysis, 6-minute walk testing, transthoracic echocardiography, cardiac MRI, cardiopulmonary exercise testing, pulmonary function testing, chest computed tomography, ventilation lung scan, and right heart catheterization with provocative maneuvers. Full details of each procedure are provided as Online Data Supplement.

**Interactions With Other NHLBI Awards and the Pulmonary Hypertension Association**

The PVDOMICS network plans to leverage the synergistic scientific and operational strengths of 2 NHLBI-awarded Investigator-Initiated Resource-Related Research Project Application (R24) grants on PH to support, facilitate, and accelerate PVDOMICS goals: (1) Pulmonary Hypertension Breakthrough Initiative and (2) the National Biological Sample and Data Repository for pulmonary arterial hypertension (PAHBiobank). Specifically, functional consequences of -omic findings can be tested in the Pulmonary Hypertension Breakthrough Initiative–derived cell lines and tissue specimens, and genetic and clinical findings in PVDOMICS could be validated in the PAHBiobank.

**Planned Analyses**

To discover molecular-based subtypes of PVD, we will apply high dimensional model-based clustering methods to the combined phenotypic and -omic variables followed by feature reduction/selection to aid in interpretability and reproducibility. We plan to partition the PVDOMICS cohort into 2 equal sized subsets, discovery and validation, cluster by phenotypes in the discovery set, determine the most informative -omic variables in a smaller subset to help maximize information content, and reduce dimensionality while minimizing costs, and then determine molecular subtypes in the full discovery set using the informative features. Full details are included in the Online Data Supplement.

**Discussion**

PVDOMICS aims to redefine classification of PVD using advanced -omic technologies. The protocol presented here has several strengths. First, there will be extensive clinical phenotyping that incorporates advanced pulmonary vascular
assessment, including provocative testing during right heart catheterization and state of the art assessment of the right ventricle. Second, -omics profiling will include constant features (genomics) and metrics that incorporate environmental and genetic influences (transcriptomics, metabolomics, proteomics, coagulomics, cell biomics). Third, carefully planned comparator groups include healthy controls and importantly, disease comparators. These disease comparators will inform differences between participants with underlying cardiopulmonary disease with and without PVD, pointing to features that promote PVD. In addition, disease comparators will facilitate analysis of pulmonary artery pressures as a continuous variable identifying PVD at an earlier time when interventions are more likely to be successful and identifying resilience features protective of PVD. Current definitions have relatively arbitrary cut offs (mean pulmonary arterial pressure: 25 versus 21–24 mmHg) with a categorical yes/no approach, ignoring the intermediate phenotypes and the granularities that exist within and between traditional WSPH PH groups.

A major strength of the PVDOMICS protocol is the inclusion of both healthy controls and disease comparators. Current WSPH classification of PH assumes clearly designated distinct clinical diagnosis among the WSPH groups 1 to 5. Often, however, clinicians note overlapping diagnoses, highlighting the possibility of patients having >1 type of PVD. PVDOMICS will enroll and deep phenotype comparator groups with matching cardiopulmonary disease and identify clinical, radiographic, and -omic features that differentiate patients with and without PH. With only healthy controls, it is difficult to understand underlying disease confounding. Furthermore, inclusion of disease comparators will facilitate analysis of pulmonary artery pressures as a continuous variable identifying PVD at an earlier time when interventions are more likely to be successful and identifying resilience features protective of PVD. Current definitions have relatively arbitrary cut offs (mean pulmonary arterial pressure: 25 versus 21–24 mmHg) with a categorical yes/no approach, ignoring the intermediate phenotypes and the granularities that exist within and between traditional WSPH PH groups.
For this protocol, a network of PVD experts developed the protocol to phenotype all forms of PVD. In addition to metrics that are standardized by several national and international groups,\textsuperscript{8–12} the working group created consensus around performance of invasive hemodynamics with provocative testing, as well as specialized imaging of the right ventricle using cardiac magnetic resonance and right ventricle–centric echocardiograms. These protocols and procedures, including information on data harmonization, are available in Online Data Supplement methods, and in more detailed fashion online (https://pvdstudy.ccf.org/pvd/) are now a resource to the community of researchers and clinicians who may wish to use these as a roadmap for future analysis of PVD.

A challenge in this phenomics study is the standardization of all data to facilitate comparison and discovery. We have written the clinical study protocols, and the Data Coordinating Center has certified competency in each protocol at all centers to address this as others have previously reported.\textsuperscript{13} Omic data collection and analysis have been similarly protocolized. In addition to standardized collection and processing procedures, we plan to batch analyze samples and use fasting samples when possible. Furthermore, we have attempted to partially overcome data variability through larger numbers and a discovery and validation approach to data analysis.

In addition to publication of results related to the main and center-specific study hypotheses, the PVDOMICS group intends to make data and resources available to the research community. To enhance the impact of the PVDOMICS Study, ancillary studies are encouraged and will be done in collaboration with study investigators during the course of the main study. Participation in, and approval of, an ancillary study is subject to review and approval of the PVDOMICS Steering Committee. At study end, PVDOMICS data and biospecimens will be archived with the NHLBI and available for future investigators to use.

Acknowledgments

We thank Christina Collart for her assistance in article preparation.

Sources of Funding

The study received grants U01 HL125218 (PI: E.B. Rosenzweig), U01 HL125205 (PI: R.P. Frantz), U01 HL125212 (PI: A.R. Hennes), U01 HL125208 (PI: F.P. Rischard), U01 HL125175 (PI: P.M. Hassoun), U01 HL125215 (PI: J.A. Leopold), and U01 HL125177 (PI: G.J. Beck).

Disclosures

A.R. Hennes has received grants from the National Institutes of Health. She has provided consultative services for Actelion, Bayer, GSK, and United Therapeutics. A. Abidov received research grant from Astellas Pharma. S.C. Erzurum chair the Pulmonary Disease American Board of Internal Medicine. R.P. Frantz is a Steering Committee member for Actelion, consultant for Bayer, consultant for DSMB, Adjudication Committee member for Lung LLC, consultant for St. Jude Medical, Steering Committee member for United Therapeutics and DSMB. N.S. Hill received research grants to institution Actelion, Bayer, Gilead, United Therapeutics, Scientific Advisory Boards, Actelion, Bayer, and Gilead. S.C. Mathai is a consultant for Actelion, Bayer, United Therapeutics Scientific Leadership Council, and Pulmonary Hypertension Association. F.P. Rischard received research funding, consulting, and advisory compensation from Bayer, Actelion, United Therapeutics, and Gilead. J.X.-J. Yuan is a member of the Steering Committee of the ENTELLIGENCE Young Investigator Program of Actelion. The other authors report no conflicts.

References


Key Words: diagnosis ■ hypertension, pulmonary ■ National Heart, Lung, and Blood Institute (U.S.) ■ pulmonary artery
PVDOMICS: A Multi-Center Study to Improve Understanding of Pulmonary Vascular Disease Through Phenomics


the PVDOMICS Study Group

Circ Res. 2017;121:1136-1139
doi: 10.1161/CIRCRESAHA.117.311737

Circulation Research is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2017 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7330. Online ISSN: 1524-4571

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circres.ahajournals.org/content/121/10/1136

Data Supplement (unedited) at:
http://circres.ahajournals.org/content/suppl/2017/10/17/CIRCRESAHA.117.311737.DC1

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation Research can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation Research is online at:
http://circres.ahajournals.org//subscriptions/
1. Study Protocol

A. PVDOMICS Organizational Structure

The PVDOMICS Study is governed by a Steering Committee composed of the Principal Investigators of the six clinical centers, the DCC, the NHLBI Project Director participating in the study, and an external Steering Committee Chair (Supplemental Figure 1). The Steering Committee has the primary responsibility for the study protocol, monitoring study conduct, and analyzing and reviewing data prior to reporting study results. Study oversight is also provided by an Observational Safety and Monitoring Board (OSMB) appointed by the NHLBI. The OSMB is responsible for monitoring patient safety and review of study performance. It consists of clinicians with expertise in pulmonary hypertension, cardiology, bioethics, and biostatistics. A NHLBI scientist other than the NHLBI’s Project Scientist serves as the Executive Secretary to the OSMB. The DCC includes central cores (heart and lung imaging and physiology cores, and clinical chemistry) that monitor the standardized collection of the study tests and procedures. There is also a Biorepository Core at the DCC that stores the serum and urine collected in order to carry out the -omic analyses.

B. Timeline
Study participants will be enrolled after informed consent over a period up to three years. Each participant’s visits will take place over the course of no more than six weeks. All participants may be contacted by telephone and/or letter at least annually after enrollment up to the end of the study (maximum of 3.5 years). Vital status and the occurrence of lung, heart, or heart-lung transplantation will be determined from this contact or by using center medical records. Cause of death will be ascertained by the site investigators.

C. Inclusion/Exclusion Criteria
The PVD patient and comparator cohorts will enroll individuals ≥18 years of age who have been referred for right heart catheterization for further evaluation of known or possible PVD and are able to perform complete diagnostic testing. Participants may be treatment naïve or with prior exposure. All subjects will sign an informed consent to perform required testing for the protocol. Participants who are dialysis dependent or are too ill to perform protocol testing will be excluded from the study. Pregnant or nursing women will also be excluded from all categories.

A control group of 100 “healthy” volunteers will also be enrolled. These individuals will be without end organ disease, ≥18 years old, with normal cardiopulmonary screening by history and exam. Specifically included subcategories include obesity, diabetes and/or hypertension without end organ disease, hyperlipidemia, or treated sleep apnea. No individuals with active malignancy other than localized non-melanoma skin cancer will be included.

D. Adjudication
Participant enrollment will be based on the primary traditional 5th World Symposium (WSPH)\(^1\). Recognizing that some patients may not fit into a single group of PH, the Adjudication Committee will evaluate the most common overlapping groups, e.g., WSPH groups 2,3 and the WSPH groups 1,3 groups. This Committee will monitor a cumulative running total for the 6 centers together in order to assure that target enrollment goals based on the primary diagnosis are met. The classic mixed phenotypes will be a predefined subset within the primary diagnosis group.

E. Broad Overview of Study Protocol
A detailed medical history, including medications, allergies, and NYHA functional class and physical exam with body composition bioimpedance measurements will be obtained. Participants will also perform a 6-minute walk test (6MWT) using standardization of encouragement phrases and patients may use oxygen if needed. 6MWT will be performed according to ATS criteria with the exceptions of including patients with pulmonary hypertension, the performance of a single test and measurement of pulse oximetry only at the beginning and end of the test \(^2\). Quality of life measures will be assessed using The Medical Outcomes Survey Short Form-36, version 2, the Minnesota Living with Heart Failure Questionnaire and a novel PAH-specific tool, the emPHasis-10 questionnaire \(^3\-5\).

Lung structure and function will be examined using 1) pulmonary function testing performed using American Thoracic Society (ATS) standardized testing guidelines \(^6\-8\), 2) computed tomography of the chest obtained as a thin section, non-contrast scan from the lung apices to the diaphragms at maximal inspiratory effort \(^9\), 3) ventilation perfusion scintigraphy performed using inhaled aerosolized technetium-99m diethylene-triamine pentacetic acid or with Xenon-133 and perfusion scanning using 99mTc
macroaggregated albumin to record the distribution of pulmonary arterial blood flow and 4) overnight sleep monitoring performed in accordance with American Academy of Sleep Medicine guidelines using a portable home multichannel device (NOXT3, Becton, Dickinson, Franklin Lakes, NJ).

Cardiac structure and function will be evaluated by 1) standard 12-lead electrocardiogram performed using a digitalized electrocardiograph machine, 2) comprehensive two-dimensional transthoracic echocardiography, including pressure wave and continuous wave Doppler analysis with image acquisition for right and left heart analysis in accordance to American Society of Echocardiography guidelines, 3) standardized Cardiac MRI exam with Gadolinium-DTPA contrast agent.

Cardiopulmonary hemodynamics will be examined by right heart catheterization with provocative testing and non-invasive or invasive cardiopulmonary exercise testing (iCPET). Right heart catheterization will be performed in the supine position with cardiac output assessed using the thermodilution technique in all participants and direct Fick technique at some centers. Following baseline assessments, participants will undergo repeat hemodynamic assessments during provocative maneuvers, including 100% O₂ (for non-CO₂ retaining subjects), inhaled NO (40 ppm, with 100% O₂) when pulmonary artery wedge pressure (PAWP) <25mmHg, and acute saline loading (500 ml intravenous pre-warmed saline over 5 minutes) when PAWP <18mmHg and right atrial pressure <15mmHg. iCPET will be performed in lieu of saline load at centers with this capability.

Cardiopulmonary exercise testing (CPET) will be done with upright cycle ergometry on all eligible subjects. In subjects who undergo iCPET, hemodynamic measures will be recorded at rest, during peak exercise, and following recovery as previously described.

F. Laboratory and -omic Collection

Blood will be collected over 2-3 visits at each center and shipped to the DCC. Blood will include resting peripheral venous samples, pulmonary artery and pulmonary capillary wedge samples obtained during right heart catheterization to obtain transpulmonary differences in -omics. If arterial access is required, arterial blood will also be obtained. The blood for RNA is collected into Paxgene, which immediately stabilizes the RNA, and the samples will be stored as described in the Paxgene protocol. DNA is collected from Buffycoats and stored in -80C. After appropriate processing and shipping, the DCC will further process and store blood for -omic analysis (see Figure 2). Planned analyses include next generation sequencing on DNA and RNA, non-targeted proteomics, coagulomics, metabolomics, and an investigation of the cell biome. In addition, the Clinical Chemistry Core at the DCC will measure standard assays such as a complete blood count, comprehensive electrolyte and metabolic panel; panels for infectious disease including HIV, hepatitis B and C; autoimmune panels (e.g., ANA, Scl-70, RF); and, studies of hemoglobinopathies.

2. Specific Protocol details

Pulmonary Function Testing

Pulmonary function testing will be performed at the time of enrollment using American Thoracic Society (ATS) standardized testing guidelines across all test sites.
Similarly performed historical test data (within 3 months) from a study site is acceptable in the event there has been no change in clinical status, or is acceptable (within 1 year) if no lung disease is evident upon completion of careful history and physical and HRCT. Pulmonary function testing includes full spirometry, both pre- and post-bronchodilator administration, measurement of lung volumes is performed in a body plethysmograph, and diffusing capacity for carbon monoxide (DLCO) is measured. Quality controls include use of both external calibration and “biological controls” at all sites for all measures, and central over-reading of both raw tracings and calculated data. A single set of prediction equations will be applied to the data centrally for final analysis, while local prediction equations at each site will substantiate the initial diagnoses. The choice to measure lung volumes in a body plethysmograph was made as we anticipate a significant proportion of WHO group 3 PAH patients to have severe chronic obstructive pulmonary disease in which gas dilution lung volume methods can markedly underestimate the lung volume due to unmeasured slowly communicating air spaces.

Six Minute Walk Test
In numerous published studies, the 6 minute walk (6MW) distance has been a key measure of disease severity and therapeutic responsiveness. Thus, our 6MW procedures, will strive to be comparable to previously published methods, with a clear eye on study subject safety. Although our protocol slightly deviates from the recently published ERS/ATS recommendations for 6MW test procedure, these ERS/ATS recommendations were largely based on experience with a population of patients with primary lung disease, rather than PAH \(^2\). In order to standardize procedures across sites for PAH patient population, we will accept those with evidence of heart failure, and for safety concerns, perform pre-titration of supplemental oxygen to a resting oxygen saturation of \(\geq 92\%\), and measure oxygen saturation at the beginning and end of the walk. Otherwise, ATS guidelines for methodology for 6MWD of using standardized encouragement phrases, standardized methods for carrying oxygen, etc. will be followed across all sites \(^2\). Primary measures will include 6 minute walk distance and Borg dyspnea score.

Sleep Monitoring
Overnight sleep monitoring will be performed in all participants upon enrollment in order to acquire information on sleep disordered breathing physiologic measures (i.e. characterize obstructive sleep apnea, central sleep apnea and periodic breathing), and to quantify extent of overnight oxygen desaturation. Similarly performed sleep studies are acceptable if performed within 1 year prior to enrollment if no substantial change in weight and no change in nighttime oxygen, or sleep disordered breathing therapy has occurred. Those participants using supplemental oxygen or positive airway pressure during sleep will be monitored on the respective baseline modality. Standardized approaches will be implemented across sites using a home sleep apnea testing system (NOX-T3, Carefusion \(\circledR\)) to record the following: nasal pressure transducer, respiratory inductance plethysmography to monitor chest and abdominal wall effort, oxygen saturation (Nonin and Wrist Ox model 3150 \(\circledR\) with Nonin Puresat \(\circledR\) oxygen saturation averaging of 3 seconds or faster for pulse rate >60 beats per minute as per American
Academy of Sleep Medicine guidelines\(^{20, 21}\), electrocardiogram (200 Hz sampling rate) to monitor cardiac rhythm abnormalities and heart rate, snoring and body position. The Noxturnal Carefusion® offline analyses program will be used to examine the quality of the recording and to assess for any urgent alerts indicating need for priority scoring. Scoring software will be used to score apneas defined as a drop in peak signal excursion by >90% of the pre-event baseline using a nasal transducer for >10 seconds in accordance with American Academy of Sleep Medicine guidelines\(^{12, 13}\). Obstructive apneas will be defined if there is persistence of thoracoabdominal effort and central apneas in the absence of thoracoabdominal effort. Hypopnea events will be scored if peak signal excursion is reduced by >30% of the pre-event baseline using nasal pressure transducer lasting for >10 seconds and associated with >3% oxygen desaturation. Periodic breathing will be defined as airflow or inductance channels increasing and decreasing at least 50% from the maximum, in a cyclic waxing and waning or "sinusoidal" manner for a consecutive period of >10 min. ECG data will be reviewed for cardiac arrhythmias including atrial and ventricular arrhythmias and atrioventricular block. Data extraction will be performed to allow for varied considerations of extent of oxygen desaturations accompanying the hypopneas. The summary data including the overall apnea hypopnea index (along with subset indices such as central apnea index and obstructive apnea hypopnea index), degree of hypoxia and periodic breathing will be statistically analyzed centrally.

**Quality of Life Questionnaires**

Traditional measures of disease severity (hemodynamics, six minute walk distance, etc) generally do not address the patient perspective and thus may fail to incorporate essential information regarding impact of disease on other aspects of life such as social and emotional experiences. These factors, along with hemodynamic and functional impairment, influence health-related quality of life (HRQOL), the “functional effect of an illness and its consequent therapy upon a patient, as perceived by the patient”. (3) Recent research has demonstrated that HRQOL is depressed in PAH\(^{4}\) and may predict clinical worsening and survival\(^ {2, 12, 13}\). Understanding the patient-centered factors that contribute to one’s experience with the disease and incorporating these into clinical assessment is of paramount importance given the high burden of disease on physical, social, and emotional function and the high burden of treatment related to administration, adverse effects, and cost\(^ {22, 23}\). Within the PVDOMICs protocol, we will assess patient-related outcomes (PROs). Traditional HRQOL tools, such as the Medical Outcomes Survey Short Form-36 and the Minnesota Living with Heart Failure Questionnaire, along with a novel PAH-specific tool, the emPHasis-10 questionnaire, will be administered to the entire study cohort\(^ {3-5}\).

**Chest CT**

For purposes of the PVDOMICS study, CT scans obtained as part of routine clinical care may be performed with or without intravenous contrast as deemed appropriate by the ordering clinician at each local center. In this setting, the CT protocol utilized will be one prescribed by the local imaging department for evaluation of the clinically suspected disease, as long as the raw data is reconstructed into 1mm or thinner sections. Furthermore, participants who have had previous CT scans performed within 1 year
prior to study enrollment and for whom the study center has reconstructed and saved the images on their picture archiving and communication system (PACS) at 1mm thickness or less, may use these previously obtained images as part of the participants full PVDOMICS imaging evaluation. CT scans performed on comparator subjects or on participants with pulmonary hypertension, but for research purposes only (not part of clinical evaluation) will be performed without intravenous contrast.

Each center has specific imaging parameters that must be implemented for the PVDOMICS chest CT scans, predetermined based on their respective hardware. These have been standardized across several manufacturer makes and models. Table 2 demonstrates an example of site specific scanner parameters for a PVDOMICS research participant. Radiation dose settings are preset according to patient’s body mass index (BMI) and are designed to minimize radiation to the participant. They range from 1.8mSv - 5.0mSv, based on size of the participant.

Scout scans will be obtained and must be performed such that when spirally scanned, the full extent of the lung is acquired at the respective lung volumes and over scanning is kept to a minimum. CT scanning must include the lungs- starting the scan slightly above the apex of the lungs and stopping it once the scan is through the base of the lungs with the field of view (FOV) fitting the lung parenchyma.

Some measurements to be collected of particular relevance to lung anatomy and morphology include quantitative assessment of main pulmonary artery and ascending aorta and qualitative analysis of emphysema, honeycombing, traction bronchiectasis, ground glass opacity, mosaic attenuation and, when applicable, thromboembolic disease.

Ventilation Perfusion Lung Scan

All patients are required to have a ventilation/perfusion (V/Q) lung scan. A V/Q scan that is performed within one year prior to enrollment may be used for collection of data in PVDOMICS. A V/Q scan performed as part of PVDOMICS will be comprised of two scintigraphic scans: one for ventilation and another for pulmonary perfusion. The ventilation scan may be performed as an Aerosol Ventilation or Gas Ventilation Scintigraphy. In Aerosol Ventilation, the bronchopulmonary distribution of an inhaled radioactive aerosol [99mTc diethylenetriamine-pentaacetic acid (DTPA)] is evaluated within the lungs. In Gas Ventilation, the pulmonary distribution of radioactive xenon (133Xe) gas is evaluated during breathing maneuvers. Pulmonary Perfusion Scintigraphy uses 99mTc macroaggregated albumin (MAA) to record the distribution of pulmonary arterial blood flow.

Electrocardiography

A standard 12-lead electrocardiogram (ECG) will be performed using a digitalized electrocardiograph machine. To ensure a standardized recording of the ECG signal from studies performed at all participating sites, the modified Goldbergwe ECG lead arrangement strategy with a 1 mV (10 mm amplitude) reference pulse will be used: 3 limb leads (leads I, II, and III), 3 augmented limb leads (aVR, aVL, and aVF), and 6 precordial leads (V1-V6). The principal objective of the ECG in this study is to establish a baseline heart rhythm, heart rate, and assess for structural abnormalities that may be relevant to the diagnosis and classification of pulmonary arterial hypertension. In particular, the following criteria will be used to diagnose right ventricular (RV)
hypertrophy: R wave in V1 >6 mm, deep S wave in V5 >10 mm, or increased R:S ratio in V1 >1.0. The following criteria will be used to diagnose right atrial enlargement: an upright P wave in lead II >2.5 mm. The diagnosis of all other abnormalities detected by ECG is in accordance to published criteria.28

**Echocardiography**

Comprehensive two-dimensional transthoracic echocardiography including pulse wave and continuous wave Doppler analysis will be performed for the complete assessment of cardiac structure, function, and geometry in all patients participating in the study. Image acquisition for right and left heart analysis will be performed in accordance to American Society of Echocardiography guidelines, including the following views obtained with the patient in the left lateral decubitus position: parasternal long and short axis, RV inflow, apical 2-, 3-, 4-, and 5 chamber, and RV focused views, suprasternal notch, and subcostal views.

Specific measurements that will be collected of particular relevance to right heart pathophysiology include: RV chamber area and fractional shortening, RV global longitudinal myocardial strain, tricuspid annular Doppler inflow velocity; of the lateral tricuspid annulus (i.e., tissue Doppler imaging); the ratio of tricuspid inflow velocity to annular relaxation velocity; and additional RV metrics including: maximal RV linear dimension at the crux of the RV, ejection time, myocardial performance index, tricuspid annular systolic plane excursion, velocity of the tricuspid regurgitant jet, and RV outflow velocity (including pulmonary insufficiency) for measurement of mean and diastolic peak gradients.

An agitated saline study using peripheral venous access will be performed in order to determine if study participants harbor an unrecognized intra-cardiac shunt, which may have important implications on pulmonary hypertension classification. Additionally, three-dimensional echocardiography will be used at selected study sites to provide additional data on ventricular geometry, quantification of RV and LV volumes, as well as ejection fraction for comparison to cardiac MR.

**Cardiac MRI**

All patients considered for participation in the PVDOMICS study will undergo a standardized Cardiac MRI exam. This dedicated protocol developed by The PVDOMICS Imaging Group and the Imaging Core Lab, is available for review on the PVDOMICS website.

The PVDOMICS Cardiac MRI protocol consists of the mandatory common part and the additional optional sequences.

The common MRI protocol will include the following:
- Assessment of cardiac structure and function (SSFP sequences);
- Hemodynamic assessment of valvular flow (phase-contrast imaging), including Qp/Qs
- Assessment of myocardial scar/fibrosis and presence of thrombus (delayed Gadolinium hyperenhancement imaging)

The optional MRI protocol will include the following:
- Assessment of diffuse fibrosis and extracellular volume fraction (using pre and post-contrast T1 mapping)
- Pulmonary artery compliance and vessel wall remodeling (using high-temporal resolution sequences)
- 4D Flow

In subjects selected with ESRD with GFR<30, or Gadolinium allergy, only non-contrast MRI will be performed. Specific MRI measurements relevant to the PAH pathophysiology include the following: structural anatomy and morphology of cardiac chambers; functional performance of the RV and the LV; anatomy and morphology of great vessels; anatomic and hemodynamic findings suggestive of an intra-cardiac shunt; presence of scar/fibrosis of the LV and/or RV.

Right Heart Catheterization with/without Exercise Testing

Definitive diagnosis of PH requires the invasive demonstration of mean pulmonary arterial (PA) pressure (mPAP) ≥25 mmHg. PA catheterization provides valuable additional information regarding pulmonary vascular phenotype such as cardiac output (CO) and left ventricular filling pressures. While the current nosology has represented the standard of care for years, recent evidence indicates that standard hemodynamics may be insufficient to optimally phenotype patients. For example, some patients with non-Group 2 PH display elevation in pulmonary capillary wedge pressure (PCWP) in response to provocative maneuvers, such as exercise or volume loading, which suggests a component of left heart disease is contributing to PH in these participants. Exercise may unmask abnormalities in PA vascular function and right ventricular (RV)-PA interaction that are not apparent at rest. Administration of inhaled nitric oxide (NO) and 100% oxygen may offer additional information or unmask occult left heart disease by augmenting preload to the left atrium.

In PVDOMICS, participants will undergo rigorous invasive hemodynamic phenotyping. Right heart catheterization will be performed in the supine position using fluid filled, balloon-tipped catheters (≥6Fr). Pressures in the right atrium, RV and PCWP will be measured both at end expiration and during spontaneous breathing at rest and then during a series of provocative maneuvers. PCWP position will be verified by fluoroscopy, the presence of typical waveforms, and O2 saturation (≥94% or <2% of systemic saturations). CO will be assessed using the thermodilution technique in all participants and direct Fick technique when available. Standard derived measures of resistance and compliance will be calculated from these raw data (see online supplement).

Following baseline assessments, participants will undergo repeat hemodynamic assessments during provocative maneuvers, including (in order) 100% O2 (for non-CO2 retaining subjects), inhaled NO (40 ppm, with 100% O2), and acute saline loading (500 ml prewarmed saline over 5 minutes). Invasive hemodynamic exercise testing will be performed in lieu of saline load at centers with this capability (see invasive CPET section).

Functional Phenotyping: The Integration of Hemodynamics with Imaging
Phenotyping across WHO groups requires characterization of patients where function and/or prognosis are defined by the pulmonary vascular component of their disease. Metrics most sensitive to pulmonary vascular physiology are needed such as pressure-volume (PV) analysis of the RV and impedance spectra analysis to accurately define RV afterload. Several centers in the PVDOMICS network will utilize state-of-the-art technology integrating cardiac MRI and/or echocardiography during or shortly after invasive assessment to acquire these metrics across WHO groups. Recent investigation has shown PV analysis to be sensitive to earlier changes than standard hemodynamics in high risk phenotypes such as scleroderma. These metrics can then be compared to non-invasive surrogates that may be used clinically.

**CPET/iCPET**

Cardiopulmonary Exercise Testing: To better assess the cardio-pulmonary function during incremental exercise, we will perform cardiopulmonary exercise test (CPET) with cycle ergometry (upright preferred) on all enrolled and eligible subjects. We will exclude based on best clinical judgement those with current exercise-induced angina, inability to cycle, exertional syncope, known potentially lethal exercise-induced arrhythmia, severe symptomatic aortic stenosis, or with resting systolic blood pressure ≥ 200 mmHg. Participants are encouraged to give a maximal effort on a standard incremental protocol (either 10W/min, 15W/min, or 20W/min) selected based on their current, self-reported physical abilities with the goal of achieving physiologic limitation at target pedal speed of 60 rpm to reach a Respiratory Exchange Ratio >1.1 and to maintain it for 30+ seconds. Actual pedal rate, heart rate, blood pressure (by arm cuff or arterial line), oxygen saturation, perceived exertion (Borg scale), and dyspnea (Borg scale) will be recorded in a worksheet, and exercise 12-lead electrocardiogram via modified 10-electrode (Mason-Likar) configuration and exhaled gases via sealed mouthpiece will be collected by validated metabolic cart and digital data transferred to core laboratory for analysis. Recovery should include a cool-down phase at a reduced work rate (e.g., unloaded pedaling) for 2 minutes followed by 3 minutes of rest with ECG and symptom monitoring. A subset of participants will undergo invasive CPET (iCPET) as part of the right heart catheterization procedure. Hemodynamic measures will be record prior to, during peak exercise, and following recovery as previously described. Systemic and pulmonary arterial blood sampling will occur at rest, peak exercise, and post-exercise. Participants undergoing an iCPET during the right heart catheterization will not undergo a non-invasive CPET.

**3. Omics Cores**

All samples will be collected based on the International Society for Biological and Environmental Repositories (ISBER) Best Practices for Repositories.

**Genomics Core.** The Genomics Core offers measure of micronuclei, Genome-wide expression profiles using RNAseq, microRNA expression and other genomic analyses including whole genome and whole exome sequencing to address the RFA goal to define endophenotypes that may reflect pre-clinical disease. The blood for RNA is collected into Paxgene, which immediately stabilizes the RNA, and the samples will be stored as described in the Paxgene protocol. The RNA will not be extracted until they
are analyzed, in order to minimize batch effects. DNA is not subject to environmental variables. The steering committee will decide which genomics will be performed.

**microRNA expression** will be used to profile 800 microRNAs across the first 25% of patient and control samples from plasma and peripheral blood mononuclear cells (PBMC) using Nanostring nCounter technology. Results will be normalized to non-human spike-in control microRNAs. The top 8 microRNAs that differentiate one or more PH subgroups, or correlate with disease severity and survival, will then be validated across all remaining samples using Taqman realtime PCR assays. Levels of micronuclei can be used to measure **DNA damage (micronucleus assay)**. DNA damage will be measured in PBMC in all patient and control groups by flow cytometry to investigate if increased levels of DNA damage may contribute to the pathogenesis of pulmonary hypertension. PBMC will be collected from Heparin vacutainer shipped to the biorepository core at room temperature overnight. **Mutation analysis of heritable PAH genes.** The PVDOMICS cohort provides the ideal platform to perform mutation analysis across all samples from patients and controls in this study. However, the constraints of the budget necessitate that mutation analysis will be confined. DNA will be extract from Buffycoats. **Genome-wide expression profiles** from PBMC will also be performed using RNAseq. Although PBMC do not completely replicate the expression profiles of the lung, this is the best available surrogate and has been used to generate distinct expression signatures in PH and other lung conditions (1-4). RNAseq is more expensive than traditional gene expression arrays but has several advantages, including the ability to identify alternative splice products that may differ between patients and controls, analyze non-coding RNAs that are not represented on most expression arrays, and provide genotyping information on coding single nucleotide polymorphisms (SNPs) that may influence gene expression levels and/or protein function. RNAseq will be performed across the first 25% of patient and control samples of PBMC RNA followed by validation of genes that differentiate one or more PH subgroups, or correlate with disease severity and survival, across all remaining samples. PBMC will be collected from Heparin vacutainer shipped to the biorepository core at room temperature overnight.

**Proteomics Core** offers both targeted and untargeted proteomic LC-MS/MS based experiments. For untargeted proteomics, samples are analyzed in an unbiased manner on either a ThermoScientific Orbitrap Elite or Fusion Lumos instrument. The analysis of plasma proteomics will involve the immunodepletion of the top 14 plasma proteins using an IgY14 LC-5 column (Sigma) prior to in-solution digestion, off-line high pH reverse phase fractionation, and LC-MS/MS analysis\(^\text{43, 44}\). The relative protein abundance can be quantified by using either a label free or an isobaric tagging method. For the label free method, the normalized peptide peak intensities are used to compare the relative abundance of proteins across the samples using the program MaxQuant\(^\text{45}\). These experiments can be performed on either the Elite or Lumos instruments. For the isobaric tagging experiments, samples are labeled with isobaric tags that allow mixing of up to 9 samples and 1 internal standard in a single experiment. These samples are labelled after digestion and mixed prior to off-line fractionation. These samples will need to be performed on the Lumos instrument. The method that will be applied to
these samples is the label free method, due to the higher number of proteins that can be identified, however, if greater quantitative accuracy is desired the isobaric tagging method can be utilized. The untargeted proteomics experiments will be performed on 100 samples and the goal of these experiments is to identify protein targets for validation.

For the targeted proteomic experiments, an LC-MS/MS method that utilizes parallel reaction monitoring (PRM) on the Fusion Lumos instrument will be utilized. PRM involves the fragmentation of two to three specific peptides from each the protein of interest. It has been shown that PRM on the Lumos instrument has similar sensitivity, linearity, and selectivity compared to multiple reaction monitoring (MRM) on a triple quadrupole instrument. In addition, PRM can be multiplexed to allow the analysis of up to 100 proteins in a single LC-MS/MS experiment. If "stable isotope-labeled" synthetic peptides are included in these experiments, the absolute abundance of each protein can be determined. The benefit of PRM is the ability to quantify multiple proteins in a single LC-MS/MS based method with good sensitivity. For proteins that are not amenable to LC-MS/MS based PRM due to poor reactivity or abundance, an Elisa based method can be used. A third option for the proteomic experiments is to analyze samples using a commercially available aptamer-based proteomic platform. The current plan is to send these samples to the Genomics and Microarray Core at the University of Colorado. The assay and data analysis will be performed by the core.

Metabolomics Core. The quantification of small molecules in biospecimens, using global untargeted (Metabolon) and specifically targeted approaches can allow for a comprehensive analysis of the changes in the metabolism and signaling pathways in a specific organ (e.g., lung) and in the whole body. BMPR2 associated PH manifests widespread metabolic changes and provides a strong rationale for the examination of the metabolome in PH (1). Measurements of the dysregulated arginine metabolome and oxidative and nitrative stress biomarkers provide a metabolic endophenotype which can be related to clinical PH phenotype. The metabolomic core will provide a state of the art analysis of plasma obtained from subjects with PAH and controls. Studies of a carefully identified study population, characterized for disease (PAH) state as well as for metabolic parameters such as anthropometry, diet, insulin resistance, and studied in the basal state and in response to challenge (dietary or exercise) are likely to provide greater insight into the pathobiology of PAH and help in identification of phenotypes and the development of novel therapeutic strategies.

Non-targeted metabolomics (METABOLON). Metabolomic profiles will be determined in plasma obtained in the study protocol/s planned by the steering committee. The global, unbiased metabolic profiling platform is based on a combination of three independent platforms: ultrahigh performance liquid chromatography/tandem mass spectrometry (UHLC/MS/MS) optimized for basic species, UHLC/MS/MS optimized for acidic species, and gas chromatography/mass spectrometry (GC/MS). This platform is described in detail in a previous publication (2,3) Targeted Metabolomics. The core has established and optimized the EICOM NOx Analyzer (ENO-20), an HPLC-based system, for measurements of nitrate and nitrite in serum, plasma and urine (4-6). The high sensitivity and specificity are accomplished with the combination of a dizao
coupling method and chromatography. The level of diazo compound is measured by absorbance at 540 nm using a visible detector. The nitrate and nitrite levels in the sample will be identified by their retention times related to the reference peaks produced by the nitrate and nitrite standards. We have an autosampler available for high throughput analyses. The Core has also the ability to measure [1] Oxidative/Nitrative Stress Amino Acid biomarkers such as Carnitine, Tyrosine, Nitrotyrosine, Dityrosine, Chlorotyrosine, M-tyrosine, O-tyrosine, Bromo-tyrosine, Phenylalanine. [2] Oxidative stress Ox-Fatty markers, HETEs, HODEs, linoleic acid, arachidonic acid, PGF-2a, F2-isoprostanes and [3] Nitrative stress ADMA/Arginine biomarkers: Arginine, Lysine, Citrulline, Homocitrulline, SDMA, ADMA, MMA. Stable isotope dilution LC/MS/MS will be used to quantify the metabolite isotopologues above from acidified plasma using positive multiple reaction monitoring which are identified using multiple pre-specified precursor product ion transitions as we have extensively published (7-9). Deuterated internal standards will be added to the samples for their respective native compounds. Increasing concentrations of analyte standards with a fixed amount of internal standard will be added to human control plasma in order to generate calibration curves for determining plasma concentrations of each analyte. Plasma for the metabolomics core is collected from EDTA blood tubes. All blood is processed within 1 hour of collection and aliquots specific for metabolomics analyses are stored in -80C.

Cell Biomics Core. This core will provide a state of art analysis of bone marrow-resident and circulating progenitor cells, and circulating endothelial cells, fibrocytes and microparticles (1-7). There is emerging evidence that abnormalities in hematopoiesis are inherent to PAH (4). The reports point towards an imbalanced differentiation in the hematopoietic system in which the subset(s) that promote vascular health is predominated by increased numbers of pathologic progenitors (2,8). Microparticles are bioactive molecules carrying cytoplasmatic vesicles generated by exocytic budding. Recently, micro flow cytometers have been available, allowing quantification and phenotyping of microparticles in an unparallel method. Early studies using conventional flow cytometers showed that increased numbers of circulating endothelial cell-derived microparticles correlated to disease severity. Guides for the processing of samples for microparticle analysis is evolving rapidly. The urine and blood samples will be collected as recommended by the Microvesicle Analysis Interest Development Group of International Society for Advancement of Cytometry. β-Adrenergic receptors (βARs) are G-protein coupled receptors that control cardiac chronotropy, inotropy, and vascular tone (9,10). Binding of β-agonists to the extracellular domain induces interaction with a heterotrimeric G protein in the intracellular compartment followed by release of α subunit, which stimulates adenylyl cyclase to produce the second messenger molecule cAMP (11). For the analyses of β-adrenergic receptor signaling in circulating cells, a 10 ml EDTA blood fix and lyse/perm method will be used 12. We have tested several fixation and permeabilization strategies and this method worked the best for pβAR and cAMP detection.

Coagulomics Core. The core will provide a comprehensive, statistically robust profile of the coagulation cascade activation, and effect of Warfarin/anti-coagulation, on ongoing thrombin activation in patients with PAH. The core will likely identify a sub-set of those
with PAH whose disease or its progression is a consequence of an underlying hypercoaguable state. We will have identified, characterized and quantified a novel hypercoaguable endophenotype, heretofore only presumed. Correlation with clinical outcome data will provide substantive support for the significance of the coagulation cascade to the development and/or progression of pulmonary hypertension. The plasma for the coagulomic measurements is collected from EDTA and Citrate tubes. All blood is processed within 1 hour of collection and aliquots are stored in -80C.

Based on steering committee approval and review of pilot study data in this patient population select coagulation system testing will be performed. We collect both EDTA (3mls) and citrated plasma (2.6 mls) which are frozen in small aliquots (100 µl) for these assays. The core has the capacity to measure plasma levels of soluble Tissue Factor (TF) antigen can be measured by ELISA and TF activity and Factor Xa is measured by chromogenic assays (from Am. Diag.) (1-4). Thrombin generation will be evaluated in plasma using multiple complementary measures of the thrombin pathway. This is optimal as clearance rates and inhibition of thrombin by naturally-occurring thrombin inhibitors in plasma are rapid. Thrombin enzyme activity can be measured by fluorogenic assay, while prothrombin fragment F1+2 (cleaved from thrombin upon activation; USCN Life Sciences), thrombin-anti-thrombin complexes, and fibrinopeptide A (cleaved from N-terminus of fibrinogen by thrombin) are measured by ELISA (Am. Diag., Seimens) (5,6). Furthermore, we can perform Thrombus Formation/Lysis assays: Briefly t-PA, u-PA, PAI-1 Ag/activities and fibrin-degradation products (D dimer) are measured by ELISA (Am. Diag. and Olman Lab), where their respective activities (where relevant) are measured by chromogenic assays (Am. Diag., Olman Lab)(7-9). Platelet β-thromboglobulin activation by ADP and TRAP can be measured by ELISA (from Am. Diag.) (10,11). vWF antigen can be measured by ELISA (Am. Diag) and PAI-1 Ag/Activity is measured as published by Dr. Olman (12). As small studies suggest that the frequency of inherited/acquired coagulation/fibrinolytic defects in this patient population is quite low and the volume of blood required exceeds our ability to safely collect it, this evaluative protocol will not be performed routinely. 

4. Planned Analyses

Primary analyses will aim to develop a new and more accurate classification of PVD. Development of molecular-based subtypes and identification of features driving the subtypes will aid in the development of a controlled vocabulary and new PVD ontologies. High-dimensional data from multiple -omic layers, e.g., genomic, transcriptomic and proteomic, will be available on all subjects, with a subset likely having more extensive data as part of the omics discovery process. While each -omic layer may be useful for subtype discovery, approaches that utilize multiple -omic and phenotypic layers, i.e., multi-omic approaches, are generally more informative and interpretable. There is an extensive and growing literature of methods and applications involving multi-omics, much of it motivated by the availability of multi-omic datasets from TCGA (The Cancer Genome Atlas). One such family of methods for integrated clustering, iCluster, iClusterPlus and iClusterBayes, has been popular in the cancer domain for discovering molecular subtypes. The general idea underlying this family is to model each layer separately (there may be many), allowing for appropriate...
parametric models depending on type of data (e.g., binary for genetic mutation data, continuous for methylation data and count for RNA sequencing data), but link them using a common set of low dimensional latent variables. By linking the -omics layers, the multiple high dimensional spaces, are reduced to a common low dimensional subspace that captures much of the total variation and can be used to cluster the samples. The most recent member of the family, iClusterBayes The most recent member of the family, iClusterBayes, has a number of additional features that make it attractive for multi-omics clustering. It provides posterior probabilities to reflect the importance of omics measures in determining the clusters (i.e., variable selection), an automatic way to determine the optimal number of clusters using the Bayesian Information Criteria, ability to incorporate prior information about omic measures and substantially increased computational efficiency compared to iClusterPlus. We intend to apply iClusterBayes to the PVDOMICS multi-omics data as our primary method to develop a new classification system and discover novel molecular subtypes of PVD. Omic measures selected by iClusterBayes' variable selection method as useful for clustering during the discovery phase of our analyses will be carried forward to the clustering validation phase as described below.

Specifically, we propose the following plan for discovering PVD types and subtypes using phenotype and molecular measurements:

1. When 750 subjects have been recruited, the first half of the anticipated study cohort, and their phenotypic variables have been measured, including demographic and clinical variables, we will compare PVD patients to their matched comparators to find phenotypic variables more likely to be related to PVD etiology than to underlying comorbidities (e.g., heart or lung disease). Variables showing differences at a false discovery rate of 10% will used to cluster subjects into novel PVD phenotype-derived groups. iClusterBayes will be used for the clustering and the Bayesian Information Criteria (BIC) will be used to select the number of clusters. Based on a preliminary analysis of 1900 subjects at the Cleveland Clinic, which have a subset of the planned PVDOMICS phenotypic variables, we expect to derive four to six phenotype groups, with the minimum group percentage being greater than 10%.

2. 50 subjects will be chosen from each phenotype group in such a way as to maximize variability across phenotype variables within each group. A full battery of -omic measures, say M variables, will be taken on biological samples from this subset, the omic discovery subject subset (ODSS), with specific choices of biological layers (e.g., RNA, DNA, metabolome) and platforms (e.g., sequencing, microarrays) to be determined.

3. To reduce the large number of -omic variables under consideration for subtype discovery, each -omic variable will be tested for differences between each pair of phenotype groups using the data from the ODSS. In addition, differences in -omic variables between PVD and matched comparators across all phenotype groups will be assessed to help eliminate measures more related to comorbidities than PVD. Power and expected false discovery rates for these comparisons assuming M=20,000 -omic measures are given in Table 5. -Omic variables will be ranked by their standardized effect sizes of differences between phenotype groups, between PVD and comparators and by their variability within phenotype groups. Ranks will be combined using rank aggregation methods to generate a single ranked list. The rank list will be pruned to
account for correlations among -omic variables. Using the filtered rank list and considerations of cost and feasibility, P1 of the M -omic variables will be selected for measurement in all first-half subjects not in the ODSS, which we label NODSS1. We anticipate NODSS1 will consist of between 300 and 500 subjects and expect P1 to be much smaller than M.

(4) After measuring the selected P1 -omic variables in the NODSS1, clustering of the P1 -omic variables within each phenotype group, i.e., a form of semi-supervised clustering, will be done using iClusterBayes to find omic-based subtypes using the combined ODSS+NODSS1 subject sets. In addition, step (3) above will be repeated using the combined subject sets and P1 -omic variables on both the phenotype groups alone and the -omic-based subtypes to validate and filter the P1 -omic variables. A smaller number of -omic variables, P2, may be chosen using the variable selection method in iClusterBayes for measurement in the remaining half of the study, labeled NODSS2, ~750 subjects.

(5) Finally, after measuring the phenotype variables and P2 -omic variables on the NODSS2 subject set, step (4) will be repeated on the combined ODSS+NODSS1+NODSS2 set and on the NODSS2 set alone. Adjusted Rand Index will be used to measure cluster stability of ODSS+NODSS1 derived types and subtypes in the NODSS2 set.

Biological validity of the derived types and subtypes will be assessed by associating cluster memberships with clinical outcomes, including survival times and event rates. We decided to wait to form phenotype-based clusters until half of the proposed study size (1500) has been enrolled to have a high probability (estimated from our preliminary data) of getting at least 50 subjects in each cluster. Our power calculations in Table 5 show that power and FDR for discriminating -omic variables between clusters fall below tolerable levels when there were less than 50 subjects in a group. In those calculations, we assumed 20,000 -omics variables were measured in Step (2). While global DNA and mRNA measurements might be considered at this step, leading to many more than 20,000 measures, we expect due to cost and information considerations to measure a focused subset of these variables potentially combined with miRNA and other lower dimensional -omics platforms (e.g., protein arrays, metabolomics, coagulomics). For example, if 10% of the 20,000 -omic variables differ between 2 groups, with 50% of these variables having a large effect, when performing testing at the 0.01 alpha level we have 76% power to detect a medium to large effect and overall FDR of 13% with 50 subjects per group compared to just 44% power and overall FDR of 21% with 25 subjects per group. Measuring all M -omic variables on more than 50 subjects per group can substantially increase power as shown in Table 5, but there may be insufficient funding to do complete measurement in more than 50 samples per cluster assuming we find 4-6 phenotype clusters. If that situation changes, either due to increased funding or decreased costs, we will consider increasing the -omic sample size per group. We propose to do multiple filtering steps on the -omic variables to help maximize the information content while minimizing costs. The extent of the filtering will depend again on funding and costs per measure.

Our primary strategy is to first cluster/group by phenotypes in a larger subject set, determine the most informative -omic variables in a subset, and then determine molecular subtypes in the larger set using the informative markers. We will also perform
cluster analyses where we cluster first with the -omic variables (in Steps (2), (4) & (5)), then either associate phenotypes with these molecular clusters or attempt to derive phenotypic subtypes within them. We will compare the phenotype-then-omic to the omic-then-phenotype approaches based on biological validity, interpretability and cluster overlap.

A weakness of the proposed iClusterBayes multi-omic clustering method is its dependence on a common cluster structure across -omic layers. The model underlying iClusterBayes assumes the same latent variables impact all layers, which does not work well when some layers have a different underlying cluster structure than others, e.g., differing number of sample clusters. To assess this possibility and add robustness to our cluster findings, we will apply the recently proposed context-dependent clustering method, Clusternomics 58, and compare the resulting clusters to those from iClusterBayes. Clusternomics uses a Bayesian Dirichlet mixture approach to model both the local cluster structure within each layer, and the global structure which arises from combinations of layer-specific clusters. An advantage of iClusterBayes over Clusternomics is its ability to select important omic drivers of the sample clusters, potentially reducing the number of omic measures that need to be assessed for accurate clustering.

Table 5 gives power and expected false discovery rates (FDR) for two-group comparisons for a mix of effect sizes, group sizes, and alpha levels assuming 20,000 variables are measured. The R package pwr was used to make the calculations. The nominal per-test alpha levels do not account for multiple testing among cluster groups but since we are primarily interested in filtering -omic features and wish to include features that may only separate two of the groups, we are less concerned about Type 1 errors. This table is useful for justifying our choice of 50 subjects to measure complete -omics in Step (2) as discussed above.

Table 6 gives minimum detectable fold changes (MDFC) between two independent groups for various number of features tested, levels of variability (as measured by a common coefficient of variation), feature intensity (as measured by average number of sequencing reads per feature) and sample sizes. A desired power level of 80% and a conservative multiple test level per feature of 0.10/# of assumed features were used in the calculations. The R package RNASeqPower, which assumed that features followed a conditional negative binomial distribution and accounted for sequencing based measurement error, was used to estimate the power of the designs. This table can be used to assess our ability to detect meaningful differences in phenotype variables at Steps (1)-(5) across phenotype clusters, assuming 50-500 phenotype variables. For example, if a set of 50 phenotype variables are tested between phenotype-derived clusters (or across a priori defined groups like WHO categories), we have 80% power to detect a fold change as small as 1.32 between groups assuming 100 subjects per group (cluster), phenotype variables are measured with very little error (Feature Intensity=High) and variability across subjects is medium (cv=0.5). This table can also be used to assess the MDFCs for -omic variables measured in Steps (2)-(5). For example, if after filtering in Step (3), we select P1=5000 -omic variables to measure in Step (4), we have a MDFC of 1.43 between two groups for say a gene expression measurement assuming 40 subjects per group (subtype), gene expression is medium and variability across subjects is small. Most of the MDFCs
in Table 6 were below 2 for low to medium variability scenarios, indicating that our
designs in these scenarios are reasonably well powered to find realistic fold changes
from human samples.

In addition to the primary goal of classification and subtype discovery, we will
attempt to build biological layer-specific and integrated-omic network representations 48,
52, e.g., co-expression or mutual information networks from transcriptome data, to aid in
understanding the biological layer complexity, degree and inter-connections. Results
from each layer will be compared and contrasted both analytically and visually. In
addition, data from each layer will be associated with clinical outcomes,
endophenotypes and newly derived type and subtypes using sparse regression
methods to gain insight into the predictability of important downstream measures.

Secondary analyses will focus on comparisons among current WHO
classifications, including comparisons to non-PH comparators and controls. Clinical and
-momic variables will be compared across existing groups. Prediction models for WHO
classes plus non-PH comparators will be built using sparse regression methods in an
effort to gain understanding of the predictability and drivers of the existing systems.

5. Data Confidentiality
Clinical centers are responsible for maintaining confidentiality of data from
subjects enrolled at their institution by following good clinical practice and their own
institutional procedures. No individual identifiers such as patient name or SSN will be
transmitted (with data or images) or stored in the DCC database. Database access will
be password protected. Data will be entered with strong encryption and the Thawte, Inc.
system will be used for secure transmission. Passwords will be changed every 75 days,
and security levels will be assigned to set privileges. The need for strict confidentiality of
all PVDOMICS records will be emphasized to clinical center and DCC staffs.
PVDOMICS paper forms and electronic media will be stored in locked cabinets. Using
fine grain access, each clinical center will have access to only its data. All modifications
to the study data will be tracked using audit tables. DCC sensitive email messages can
be sent (using the word “confidential”) through a secure Web-based system.

The DCC’s internal policies and Standard Operating Procedures are designed to
ensure adherence to Good Clinical Practice, the ICH Harmonized Tripartite Guidelines
dealing with human subject research. The DCC is fully compliant with HIPAA Privacy
and Security Rules, including the HiTECH Act.

At the end of the study the DCC will prepare a clean master data set including
analysis files and study documentation (e.g., forms, manuals, protocol) for archival at
NHBLI. Also the de-identified data files and unused biospecimens will be submitted to
the NHLBI Biological Specimen and Data Repository Information Coordinating Center
(BioLINCC) Repository for sharing. All samples and data transferred to the Repository
will be under the custodianship of the NHLBI.

Regarding the information obtained from –omics analyses, the biobank consent
form states:
Your privacy and the confidentiality of biological samples are very important to us, and we
will make every effort to protect them, including:
1) New identification numbers (codes) will be assigned to your biological samples prior to going into the biobank. This coding will stop any linking of the samples with your identity. The PVDOMICS Data Coordinating Center at the Cleveland Clinic in Cleveland, Ohio will be responsible for creating the biobank identification codes and organizing the samples going into the biobank. The Data Coordinating Center does not have access to any of your personal information, such as your name, address, phone number, social security number, or e-mail address. There is no link of the biobank codes to your name or other personal information. Through this process, the data and biological samples will be de-identified.

2) Biological samples are only available to researchers under terms and conditions in line with this informed consent document and NHLBI policies. Each request is reviewed in detail. Researchers who are approved to get biological samples are expected to take data protection measures and to respect the privacy of research participants.

3) A federal law, called the Genetic Information Nondiscrimination Act (GINA), effective May 21, 2010, generally makes it illegal for health insurance companies, group health plans, and most employers to discriminate against you based on genetic information. This law will protect you in the following ways: (1) Health insurance companies and group health plans may not request your genetic information that we get from this research. (2) Health insurance companies and group health plans may not use your genetic information when making decisions regarding your eligibility or premiums. (3) Employers with 15 or more employees may not use your genetic information from this research when making a decision to hire, promote, or fire you or when setting terms of your employment. This federal law does not apply to companies that sell life insurance, disability insurance, or long-term care insurance and does not protect you against use of this information for an already diagnosed genetic condition or disease.
<table>
<thead>
<tr>
<th>Center</th>
<th>Specific Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brigham and Women’s Hospital</td>
<td>Patients with exercise-induced PH will have -omic features that are similar to those with Group 1 PH, different from those with normal exercise hemodynamics. This will be distinguishable based on endothelial -omic pathophenotypes.</td>
</tr>
<tr>
<td>Columbia University/Cornell University</td>
<td>Patients with parenchymal lung disease and moderate to severe PH will have -omic signatures that reflect not only underlying pathogenesis (emphysema or fibrosis) but will have patterns of genomic vulnerability similar to patients with Group 1 PH.</td>
</tr>
<tr>
<td>Johns Hopkins University</td>
<td>Connective tissue disease patients with PH (largely systemic sclerosis) will have endothelial dysfunction measured as abnormalities in nitric oxide production, arginine/ornithine/citrulline metabolism and vasoactive mediators that are similar to patients with Group 1 PH and different from scleroderma patients without PH.</td>
</tr>
<tr>
<td>Mayo Clinics</td>
<td>Genomic, transcriptomic and metabolomic patterns will distinguish the degree of right ventricular compensation for a similar degree of right ventricular afterload in patients with pulmonary hypertension, regardless of the underlying WHO category.</td>
</tr>
<tr>
<td>University of Arizona</td>
<td>Racial and ethnic (or ancestry) differences in transcriptomics, epigenetics and mitochondrial haplotypes will inform PH pathogenesis by similarities and differences among similar phenotypes with PH.</td>
</tr>
<tr>
<td>Vanderbilt University Medical Center</td>
<td>Combined Group 1 and 2 PH occurs in a context of predisposing genetic variants and metabolic features that</td>
</tr>
<tr>
<td> </td>
<td>promote pulmonary vascular disease. These predisposing features can be used to identify endophenotypes and dynamically identify subsets of PH that are likely to respond to therapeutics.</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>WSPH Groups</th>
<th>Hemodynamic Definition</th>
<th>Subtypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (n=300)</td>
<td>rest mPAP $\geq 25$ AND PVR $&gt; 3.0$</td>
<td>PAH, heritable PAH, or PAH associated with HIV, portopulmonary hypertension, drugs and toxins, CHD, schistosomiasis, CTD*, PVOD or PCH</td>
</tr>
<tr>
<td>2 (n=300)</td>
<td>mPAP $\geq 25$ AND PCW $&gt; 15$ Hemodynamics will be that of combined precapillary and postcapillary (Cpc-PH) mPAP $\geq 25$ with either: PVR $&gt; 3.0$ if CO $&gt; 4$ or if CO $&lt; 4$, then DPG $&gt; 7$ or TPG $&gt; 12$</td>
<td>HFrEF, HFpEF, VHD, RCM, HCM</td>
</tr>
<tr>
<td>3 (n=300)</td>
<td>mPAP $\geq 25$ mmHg and will be further categorized as follows as per the WSPH: Moderate PH-COPD, PH-IPF and PH-CPFE (35$&gt; mPAP &gt; 25$); or Severe PH-COPD, severe PH-IPF, severe PH-CPFE defined as mPAP $\geq 35$ or $\geq 25$ with CI $\leq 2.0$ 0 L/min/m2. Other COPD/IPF/CPFE defined as in Supplemental Data, Appendix 1. OSA, OHS, other ILD.</td>
<td></td>
</tr>
</tbody>
</table>
ILD, non-parenchymal restrictive lung disease (RLD; neuromuscular disease or thoracic cage abnormalities), OSA, OHS (Supplemental Data, Appendix 1) will similarly be hemodynamically categorized.

| 4 (n=50) | mPAP ≥ 25 AND mPCW ≤ 15 mmHg | Patients with high probability V/Q or low/intermediate V/Q and (+) CTA or pulmonary angiogram consistent with chronic thromboembolic (CTE) disease, supportive pulmonary angiogram and > 3 months therapeutic anticoagulation |
| 5 (n=50) | mPAP ≥ 25 AND mPCW ≤ 15 mmHg | sarcoidosis (Supplemental Data, Appendix 1), myeloproliferative disease, or hemoglobinopathy (Sickle cell or Thalassemia) |

mPAP = mean pulmonary arterial pressure, PVR = pulmonary vascular resistance, PCW = pulmonary capillary wedge pressure, PAH = pulmonary arterial hypertension, CHD = congenital heart disease, CTD=connective tissue disease, PVOD=pulmonary veno-occlusive disease, PCH=pulmonary capillary hemangiomatosis. HFrEF = heart
failure with reduced ejection fraction, HFpEF = heart failure with preserved ejection fraction, VHD = valvular heart disease, RCM = restrictive cardiomyopathy, HCM = hypertrophic cardiomyopathy, COPD = chronic obstructive pulmonary disease, IPF = idiopathic pulmonary fibrosis, CPFE = combined pulmonary fibrosis and emphysema, OSA = obstructive sleep apnea, OHS = obesity hypoventilation syndrome, ILD = interstitial lung disease. *includes scleroderma, systemic lupus erythematosus, mixed connective tissue disease, Sjogren’s syndrome, rheumatoid arthritis.
### Online Table III. Pulmonary Vascular Disease Comparator and Healthy Control Enrollment

<table>
<thead>
<tr>
<th>WSPH Group for Comparison</th>
<th>Hemodynamics</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group 1 (n=120)</strong></td>
<td>1) borderline PH: mPAP 21 to &lt; 25mmHg; or 2) exercise induced (ePAH) with normal resting hemodynamics but exercise mPAP &gt; 30mmHg, flow (Q) &lt; 10 L/min and mPAP-Q slope &gt; 3 (mmHg*min/L)</td>
<td>mPAP&lt;25mmHg, exercise-induced PH, blood relatives of HPAH, CTD with mPAP&lt;25mmHg</td>
</tr>
<tr>
<td><strong>Group 2 (n=125)</strong></td>
<td>Mild PVD risk associated with LHD is defined as mPAP &lt; 25. Moderate PVD risk is defined as Isolated post-capillary pulmonary hypertension (Ipc-PH) mPAP ≥25mmHg, PVR &lt; 3 Wood Units,</td>
<td>50% moderate PVD and 50% mild to no PVD, HFrEF, HFpEF, RCM, HCM, VHD and exclude: Takotsubo cardiomyopathy, apical ballooning, acute myocarditis, &lt; 6 months post-surgical or catheter based valvular intervention.</td>
</tr>
<tr>
<td>Group 3 (n=125)</td>
<td>diastolic pressure gradient (DPG) &lt; 7mmHg or resting mPCW &lt; 15mmHg with provokable mPCW &gt; 18mmHg or mPCW &gt; 15mmHg with large v waves following challenge</td>
<td>1. Mild to no associated PVD risk (mPAP &lt; 21mmHg. 2. Moderate associated PVD (mPAP 21 to &lt; 25mmHg)</td>
</tr>
<tr>
<td>----------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Group 4 (n=30)</td>
<td>mPAP&lt;25mmHg</td>
<td>Chronic PE without associated PH</td>
</tr>
<tr>
<td>Healthy Control (n=100)</td>
<td>mPAP &lt;25mmHg</td>
<td>Partners or spouses or accompanying friends of patients participating in the study will be recruited to serve as true healthy controls. An effort will be made to maintain parity with regard to race, ethnicity, age and BMI with the patient population being recruited.</td>
</tr>
</tbody>
</table>

mPAP = mean pulmonary arterial pressure, PVR = pulmonary vascular resistance, PCW = pulmonary capillary wedge pressure, PAH = pulmonary arterial hypertension, CHD = congenital heart disease, CTD=connective tissue disease, PVOD=pulmonary veno-occlusive disease, PCH=pulmonary capillary hemangiomatosis. HFrEF = heart failure with reduced ejection fraction, HFpEF = heart failure with preserved ejection fraction, VHD = valvular heart disease, RCM = restrictive cardiomyopathy, HCM = hypertrophic cardiomyopathy, COPD = chronic obstructive pulmonary disease, IPF = idiopathic pulmonary fibrosis, CPFE = combined pulmonary fibrosis and emphysema, OSA = obstructive sleep apnea, OHS = obesity hypoventilation syndrome, ILD = interstitial lung disease. *includes scleroderma, systemic lupus erythematosus, mixed connective tissue disease, Sjogren’s syndrome, rheumatoid arthritis.
Online Table IV. – Center-Specific Scanner Parameters

<table>
<thead>
<tr>
<th>Scanner Parameters</th>
<th>Institution X</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scanner Make</strong></td>
<td>Siemens</td>
</tr>
<tr>
<td><strong>Scanner Model</strong></td>
<td>Somatom Dual Source</td>
</tr>
<tr>
<td><strong>Scan Type</strong></td>
<td>Spiral</td>
</tr>
<tr>
<td><strong>Detector Configuration</strong></td>
<td>64 x 0.6</td>
</tr>
<tr>
<td><strong>Rotation time (s)</strong></td>
<td>~ 0.5</td>
</tr>
<tr>
<td><strong>kV</strong></td>
<td>120</td>
</tr>
<tr>
<td><strong>mA/mAs</strong></td>
<td>BMI based</td>
</tr>
<tr>
<td><strong>Pitch</strong></td>
<td>1.0</td>
</tr>
<tr>
<td><strong>Reconstruction Algorithm Standard</strong></td>
<td>B35</td>
</tr>
<tr>
<td><strong>Reconstruction Algorithm Lung</strong></td>
<td>B60</td>
</tr>
<tr>
<td><strong>Additional Image Filters</strong></td>
<td>None</td>
</tr>
<tr>
<td><strong>Iterative Reconstruction</strong></td>
<td>Not to be used</td>
</tr>
<tr>
<td><strong>Dose modulation</strong></td>
<td>Off</td>
</tr>
<tr>
<td><strong>Thickness (mm)</strong></td>
<td>1.0</td>
</tr>
<tr>
<td><strong>Interval (mm)</strong></td>
<td>0.8</td>
</tr>
</tbody>
</table>
Online Table V. Average power and expected false discovery rate for two-group t-tests by percentage of true effects, size of true effects, alpha-level per test and sample size per group assuming 20,000 tested variables and the use of two-sample t-tests per variable with pooled standard deviation.

<table>
<thead>
<tr>
<th>% of Assoc Vars</th>
<th>Effect Size Dist</th>
<th>Alpha Per Test</th>
<th>FD R (%)</th>
<th>Ave Power (%)</th>
<th>Med/Large Power (%)</th>
<th>FD R (%)</th>
<th>Ave Power (%)</th>
<th>Med/Large Power (%)</th>
<th>FD R (%)</th>
<th>Ave Power (%)</th>
<th>Med/Large Power (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>25/50/25</td>
<td>0.01</td>
<td>44</td>
<td>24</td>
<td>32</td>
<td>29</td>
<td>47</td>
<td>61</td>
<td>24</td>
<td>61</td>
<td>78</td>
</tr>
<tr>
<td>5</td>
<td>25/50/25</td>
<td>0.05</td>
<td>69</td>
<td>43</td>
<td>54</td>
<td>60</td>
<td>63</td>
<td>79</td>
<td>56</td>
<td>74</td>
<td>91</td>
</tr>
<tr>
<td>5</td>
<td>25/50/25</td>
<td>0.10</td>
<td>78</td>
<td>53</td>
<td>65</td>
<td>73</td>
<td>71</td>
<td>86</td>
<td>71</td>
<td>79</td>
<td>95</td>
</tr>
<tr>
<td>5</td>
<td>25/25/50</td>
<td>0.01</td>
<td>36</td>
<td>34</td>
<td>44</td>
<td>25</td>
<td>58</td>
<td>76</td>
<td>22</td>
<td>68</td>
<td>88</td>
</tr>
<tr>
<td>5</td>
<td>25/25/50</td>
<td>0.05</td>
<td>64</td>
<td>52</td>
<td>66</td>
<td>57</td>
<td>70</td>
<td>88</td>
<td>55</td>
<td>77</td>
<td>95</td>
</tr>
<tr>
<td>5</td>
<td>25/25/50</td>
<td>0.10</td>
<td>78</td>
<td>53</td>
<td>65</td>
<td>73</td>
<td>71</td>
<td>86</td>
<td>70</td>
<td>81</td>
<td>97</td>
</tr>
<tr>
<td>10</td>
<td>25/50/25</td>
<td>0.01</td>
<td>27</td>
<td>24</td>
<td>32</td>
<td>16</td>
<td>47</td>
<td>61</td>
<td>13</td>
<td>61</td>
<td>78</td>
</tr>
<tr>
<td>10</td>
<td>25/50/25</td>
<td>0.05</td>
<td>51</td>
<td>43</td>
<td>54</td>
<td>42</td>
<td>63</td>
<td>79</td>
<td>38</td>
<td>74</td>
<td>91</td>
</tr>
<tr>
<td>10</td>
<td>25/50/25</td>
<td>0.10</td>
<td>78</td>
<td>53</td>
<td>65</td>
<td>73</td>
<td>71</td>
<td>86</td>
<td>71</td>
<td>79</td>
<td>95</td>
</tr>
<tr>
<td>10</td>
<td>25/25/50</td>
<td>0.01</td>
<td>21</td>
<td>34</td>
<td>44</td>
<td>13</td>
<td>58</td>
<td>76</td>
<td>22</td>
<td>68</td>
<td>88</td>
</tr>
<tr>
<td>10</td>
<td>25/25/50</td>
<td>0.05</td>
<td>46</td>
<td>52</td>
<td>66</td>
<td>39</td>
<td>70</td>
<td>88</td>
<td>37</td>
<td>77</td>
<td>95</td>
</tr>
<tr>
<td>10</td>
<td>25/25/50</td>
<td>0.10</td>
<td>59</td>
<td>62</td>
<td>76</td>
<td>54</td>
<td>76</td>
<td>93</td>
<td>53</td>
<td>81</td>
<td>97</td>
</tr>
</tbody>
</table>

% of Assoc Vars = percentage of tested variables with non-zero effect. Effect Size Dist = percentage of non-zero effects in each of 3 Cohen effect size groups: small, medium & large, given in the format small.
Alpha Per Test = type-I error rate for 2-sample t-test per variable. FDR = expected percentage of tested variables that are rejected at the specified type-I error rate Alpha but have zero effect among all rejected tests. Ave Power = power to reject a null hypothesis when the variable has a non-zero effect averaged across the entire effect size distribution. Med/High Power = power to reject a null hypothesis when the variable has a non-zero effect average across the medium and large effect size groups.
Online Table VI. Minimum detectable fold change by subgroup size, feature intensity, feature variability and number of features assuming desired 80% power, alpha level per test of 0.10/# of features.

<table>
<thead>
<tr>
<th>Feature Intensity</th>
<th># Features</th>
<th>Variability</th>
<th>Large Subgroups: 100 vs 100</th>
<th>Medium Subgroups: 40 vs 40</th>
<th>Small Subgroups: 10 vs 10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Low: cv=0.3</td>
<td>20000</td>
<td>1.40</td>
<td>1.27</td>
<td>1.26</td>
<td>1.69</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.69</td>
<td>1.47</td>
<td>1.44</td>
<td>2.87</td>
</tr>
<tr>
<td>Medium: cv=0.5</td>
<td>20000</td>
<td>1.57</td>
<td>1.48</td>
<td>1.47</td>
<td>2.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.04</td>
<td>1.85</td>
<td>1.83</td>
<td>4.18</td>
</tr>
<tr>
<td>High: cv=1</td>
<td>20000</td>
<td>2.23</td>
<td>2.16</td>
<td>2.15</td>
<td>3.55</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.55</td>
<td>3.37</td>
<td>3.35</td>
<td>12.63</td>
</tr>
<tr>
<td>Low: cv=0.3</td>
<td>5000</td>
<td>1.37</td>
<td>1.26</td>
<td>1.24</td>
<td>1.64</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.64</td>
<td>1.43</td>
<td>1.41</td>
<td>2.71</td>
</tr>
<tr>
<td>Medium: cv=0.5</td>
<td>5000</td>
<td>1.53</td>
<td>1.45</td>
<td>1.44</td>
<td>1.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.97</td>
<td>1.79</td>
<td>1.77</td>
<td>3.86</td>
</tr>
<tr>
<td>High: cv=1</td>
<td>5000</td>
<td>2.13</td>
<td>2.07</td>
<td>2.06</td>
<td>3.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.31</td>
<td>3.15</td>
<td>3.13</td>
<td>10.97</td>
</tr>
<tr>
<td>Low: cv=0.3</td>
<td>500</td>
<td>1.32</td>
<td>1.23</td>
<td>1.21</td>
<td>1.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.56</td>
<td>1.38</td>
<td>1.36</td>
<td>2.43</td>
</tr>
<tr>
<td>Medium: cv=0.5</td>
<td>500</td>
<td>1.46</td>
<td>1.39</td>
<td>1.38</td>
<td>1.83</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.83</td>
<td>1.68</td>
<td>1.67</td>
<td>3.34</td>
</tr>
<tr>
<td>High: cv=1</td>
<td>500</td>
<td>1.97</td>
<td>1.91</td>
<td>1.91</td>
<td>2.91</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.91</td>
<td>2.79</td>
<td>2.77</td>
<td>8.49</td>
</tr>
<tr>
<td>Low: cv=0.3</td>
<td>50</td>
<td>1.27</td>
<td>1.19</td>
<td>1.18</td>
<td>1.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.47</td>
<td>1.32</td>
<td>1.30</td>
<td>2.15</td>
</tr>
<tr>
<td>Medium: cv=0.5</td>
<td>50</td>
<td>1.39</td>
<td>1.33</td>
<td>1.32</td>
<td>1.68</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.68</td>
<td>1.57</td>
<td>1.55</td>
<td>2.83</td>
</tr>
<tr>
<td>High: cv=1</td>
<td>50</td>
<td>1.79</td>
<td>1.75</td>
<td>1.74</td>
<td>2.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.51</td>
<td>2.42</td>
<td>2.41</td>
<td>6.32</td>
</tr>
</tbody>
</table>

Feature Intensity: Low=10 reads per feature, Medium=100 reads per feature
Variability cv=coefficient of variation

30
High=1000 reads per feature
APPENDIX 1: Definitions of Parenchymal and Non-Parenchymal Lung Diseases

Diagnosis of COPD 59

<table>
<thead>
<tr>
<th>Severity</th>
<th>Postbronchodilator</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FEV1/FVC</td>
</tr>
<tr>
<td>At risk*</td>
<td>&gt; 0.7</td>
</tr>
<tr>
<td>Mild COPD</td>
<td>≤ 0.7</td>
</tr>
<tr>
<td>Moderate COPD</td>
<td>≤ 0.7</td>
</tr>
<tr>
<td>Severe COPD</td>
<td>≤ 0.7</td>
</tr>
<tr>
<td>Very severe COPD</td>
<td>≤ 0.7</td>
</tr>
</tbody>
</table>

*At risk: history of or current use of inhaled tobacco

Diagnosis of IPF 60

a. Exclusion of other known causes of ILD (e.g. domestic and occupational environmental exposures, connective tissue disease, and drug toxicity)
b. The presence of a UIP pattern on HRCT in patients not subjected to surgical lung biopsy
c. Specific combinations of HRCT and surgical lung biopsy pattern in patients subjected to surgical lung biopsy (see Tables A1.1, A1.2, A1.3)

Severity of IPF

- Mild FVC > 70% and DLCO > 60%
- Moderate FVC 50-70% or DLCO 40 to 60% (if FVC and DLCO are discordant, more severely reduced value determines overall severity)
- Severe FVC <50% or DLCO < 40%

Diagnosis of CPFE 61

Radiographic evidence of centrilobular and/or paraseptal emphysema in the upper lobes and pulmonary fibrosis in the lower lobes

Diagnosis of SSc-ILD 62

Pulmonary fibrosis seen on high resolution CT or chest radiography, most pronounced in the basilar portions of the lungs, or occurrence of “velcro” crackles on auscultation, not due to another cause such as congestive heart failure.

Diagnosis of OSA 63

Number of obstructive events (apneas, hypopneas + respiratory event related arousals) on PSG is greater than 15 events/hr or greater than 5/hour in a patient who reports any of the following:

- Daytime sleepiness
- Unrefreshing sleep
- Fatigue
- Insomnia
- Waking up breath holding, gasping, or choking
- Bed partner describing loud snoring, breath interruptions, or both
OSA severity:
Mild for RDI ≥ 5 and < 15
Moderate for RDI ≥ 15 and < 30
Severe for RDI > 30/hr

Diagnosis of OHS  
Sleep disordered breathing (either OSA or central sleep apnea or both)
BMI > 30 kg/m2
Daytime hypercapnia (PCO2 > 45 mmHg) in the absence of other known causes of alveolar hypoventilation

Diagnosis of Sarcoidosis  
Histologic findings of noncaseating epitheliod cell granulomas on biopsy and exclusion of other causes
Online Figure I. PVDOMICS Table of Organization
REFERENCES


