Early View

Study protocol

Rationale and design of the prognostic transcriptomic signature in fibrotic hypersensitivity pneumonitis (PREDICT) study

Evans R. Fernández Pérez, Sonia M. Leach, Brian Vestal

Please cite this article as: Fernández Pérez ER, Leach SM, Vestal B. Rationale and design of the prognostic transcriptomic signature in fibrotic hypersensitivity pneumonitis (PREDICT) study. ERJ Open Res 2023; in press (https://doi.org/10.1183/23120541.00625-2023).

This manuscript has recently been accepted for publication in the ERJ Open Research. It is published here in its accepted form prior to copyediting and typesetting by our production team. After these production processes are complete and the authors have approved the resulting proofs, the article will move to the latest issue of the ERJOR online.

Copyright ©The authors 2023. This version is distributed under the terms of the Creative Commons Attribution Non-Commercial Licence 4.0. For commercial reproduction rights and permissions contact permissions@ersnet.org
Rationale and Design of the Prognostic Transcriptomic Signature in Fibrotic Hypersensitivity Pneumonitis (PREDICT) Study

Evans R. Fernández Pérez, MD, MS;¹ ORCID: 0000-0001-7535-1591
Sonia M. Leach, PhD;² ORCID: 0000-0003-2910-069X
Brian Vestal, PhD;² ORCID: 0000-0002-3772-1691

Institutional Affiliations:
1. Division of Pulmonary, Critical Care and Sleep Medicine, National Jewish Health, Denver, CO, USA.
2. Center for Genes, Environment and Health, National Jewish Health, Denver, Colorado, USA.

Correspondence to:
Evans R. Fernández Pérez, M.D., M.S.
Division of Pulmonary, Critical Care and Sleep Medicine
National Jewish Health; Southside Building, Office G12
1400 Jackson Street, Denver, CO 80206
E-mail: Fernandevanes@njhealth.org

Trial Funding: NIH/NHLBI R01HL148437.
Article type: Clinical Study Design.

Contribution: All authors (ERFP, SML, VB) contributed to the manuscript from the outset, and read and approved the final draft. All authors are the guarantor of the paper and take responsibility for the work's integrity as a whole from inception to published article.

Running head: PREDICT Study Protocol.

Key words: hypersensitivity pneumonitis, gene expression

Word count: 3271
ABSTRACT

Hypersensitivity pneumonitis (HP) is an immunologically mediated form of lung disease, resulting from inhalational exposure to a large variety of antigens. A subgroup of patients with fibrotic HP develops symptomatic, functional, and radiographic disease progression. Mortality occurs primarily from respiratory failure as a result of progressive and self-sustaining lung injury that often occurs despite immunosuppression and removal of the inciting antigen.

The development and validation of a prognostic transcriptomic signature for fibrotic HP (PREDICT-HP) is an observational multicenter cohort study designed to explore a transcriptomic signature from peripheral blood mononuclear cells in patients with fibrotic HP that is predictive of disease progression.

This article describes the design and rationale for the PREDICT-HP study.

This study will enroll approximately 135 patients with fibrotic HP at approximately seven academic medical sites.

Participants with a confirmed diagnosis of fibrotic HP are followed over 24 months and undergo physical examinations, self-administered questionnaires, chest computed tomography, pulmonary function tests, 6-minute walk and blood testing for transcriptomic analyses. At each 6-month follow-up visit the study will assess the participant’s clinical course and clinical events including hospitalizations and respiratory exacerbations.

The PREDICT study has the potential to enhance our ability to predict disease progression and fundamentally advance our understanding of the pathobiology of FHP disease progression.
INTRODUCTION

HP is a complex lung disease that occurs in genetically susceptible individuals previously sensitized to a variety of inhaled inciting antigens. The estimated yearly prevalence of HP in the USA ranges from 1.67 to 2.71 cases per 100,000 persons.\(^1\)

While some patients with fibrotic HP (FHP) remain clinically stable, a subset develops symptomatic, functional and radiographic disease progression that results in disability, death, or lung transplant and often occurs despite pharmacological therapy and removal of the inciting antigen.\(^2\)-\(^4\)

Gene expression profiling has the ability to provide a window into the molecular biological states that are unique to well-defined circumstances, such as the disease course,\(^5\)-\(^8\) and may complement our current clinical assessment of phenotyping FHP.\(^9\)

Gene expression patterns associated with clinical, radiographic, and pathological features of HP at baseline have been identified.\(^10\) However, the current traditional clinical assessment does not include a validated prognostic transcriptomic signature prototypical of FHP progression which could prove useful in risk stratification and outcome prediction in FHP.

The purpose of this article is to describe the design of and rationale for the development and validation of a prognostic transcriptomic signature for FHP (PREDICT-HP). The National Heart, Lung, and Blood Institute supports this prospective multicenter study. It involves a collaborative effort of seven clinical centers and aims to determine if a risk indicative transcriptomic signature in peripheral blood mononuclear cells (low-risk and accessible alternative to bronchoscopy or lung biopsy) at baseline and over time used separately or in combination with traditional clinical data can predict FHP progression.
METHODS

Objectives

We hypothesize that risk indicative, gene expression signatures characterize the FHP clinical course, which can then be used to refine the prognosis and predict the clinical outcome of this complex disease. We investigate our hypotheses through the following objectives:

Primary objective: To establish transcriptomic signatures at time of initial presentation in peripheral blood mononuclear cells from patients with fibrotic HP in a discovery cohort and then validate the signature in an independent FHP replication cohort.

Secondary objective: To establish a time-course transcriptomic profile predictive of disease progression in the discovery cohort and validate the signature profile in the independent FHP replication cohort.

Study Design and Participants

This is a multicenter prospective observational study with two parallel independent study populations with FHP. The study will target 135 patients for enrollment at seven clinical centers in the USA. The study design schema is shown in Figure 1. Participants are followed every six months for 24 months. The relevant Institutional Review Boards approved the protocol before patient enrollment at each clinical center. The study is conducted in accordance with the protocol, the ethical principles of the Declaration of Helsinki and Council for International Organizations of Medical Sciences International Ethical Guidelines, the International Conference on Harmonization Good Clinical Practice guidelines, and applicable local laws and regulations. The study is registered on ClinicalTrials.gov NCT04844359.
Eligibility Criteria and Exclusions

Physicians at the enrolling sites are instructed to approach every eligible patient receiving care at their ILD clinic about participation in the study. Eligible patients are aged 18 to 80 years who meet the initial entry criteria and are diagnosed with FHP or with a provisional high-confidence diagnosis based on diagnostic guideline recommendations and integration of the clinical, imaging, bronchoscopic and, when available, surgical lung biopsy data (Table 1). Enrolled patients will receive usual care for HP as defined by the treating physician.

Assessments

At enrollment, eligible participants undergo physical examinations and standardized history focused on past medical, surgical history (e.g., including diagnostic testing such as bronchoalveolar lavage, surgical lung biopsy), medications, social and familial history. Detailed information on occupational, recreational, environmental, and residential exposures through questionnaires is obtained to determine an identifiable, indeterminate or unidentifiable inciting antigen exposure. At each visit, the exposure history is reassessed and documented to determine the effect of an abated or ongoing inciting antigen exposure. The clinical, imaging, and histopathological findings are verified centrally to ensure the correct phenotype for study enrollment.

At enrollment, data on clinical evidence of FHP progression within 12, 12-24, 24-48, and >48 months is documented based on either:

- 5–10% relative decline in FVC%.
- 5–10% relative decline in FVC% and worsening symptoms.
- >10% relative decline in FVC%.
- Increase extent of fibrosis on pre-screening HRCT.
- Other measures of clinical worsening attributed to progressive lung disease (e.g., increased oxygen requirement, decreased diffusion capacity).

Scheduled assessments are shown in the Schedule of Activities in Table 2.

**Questionnaires:** Every six months, patients self-administered patient-reported outcome measure questionnaires (the University of California at San Diego Shortness-of-Breath questionnaire [UCSD]\(^\text{12}\) and the St. George’s Respiratory questionnaire [SGRQ]\(^\text{13}\)) are collected to assess dyspnea, fatigue, and quality of life.

**Physiology:** Every six months, patients undergo pre-bronchodilator pulmonary function testing, six-minute walk test distance and Borg scale. The American Thoracic Society/European Respiratory Society guidelines serve as the primary guidance for the conduct of spirometric,\(^\text{14}\) lung volume\(^\text{15}\) and the single-breath carbon monoxide diffusing capacity (DLCO) measurements.\(^\text{16}\) DLCO values are adjusted for the most recent hemoglobin value.

**Radiology:** HRCT scans are obtained at baseline, 12 months and 24 months and then are reviewed by an expert radiologist blinded to identifiers, date and clinical data. FHP CT morphologic patterns are classified on a three-point scale based on level of confidence (typical, compatible, or indeterminate).\(^\text{11}\) The presence, extent, and distribution of individual CT features are classified using standard definitions\(^\text{17}\) and scored visually as a percent of lung involved to the nearest 5%. We will quantify HRCT lung fibrosis extent using the data-driven texture analysis (DTA) method over 52 weeks. In prior work, we have described DTA, a computer method that applies deep learning for automatic lung fibrosis detection and quantitation (DTA score).\(^\text{18-20}\)
Histopathology: Given the prognostic importance of baseline histopathology and the significance of including well-phenotype FHP patients, for every study subject with available surgical lung biopsy specimens, using the digital images, an expert blinded pathologist will analyze and score the slides for the traditional histologic features of HP using guideline criteria.\textsuperscript{11}

Biospecimen collection: Blood is collected at baseline, 12 months and 24 months, by phlebotomy, with aliquots sent for peripheral blood mononuclear cells (PBMC). Aliquots of whole blood are placed in solutions designed to preserve RNA and DNA integrity. Also, whole blood samples are collected in PAXgene blood RNA tubes. PBMCs are separated using a single-tube separation system, and serum and plasma are stored frozen. In addition to the study’s primary objective, the informed consent is written to allow for future integrated multiomics research using collected samples and sharing of deidentified data within the research community.

Clinical Measures of Disease Progression

- Rate of decline in FVC (mL) over 24 months.
- Absolute change from baseline in FVC (mL and percent predicted), and DLCO percent predicted at months 6, 12, 18 and 24.
- Proportion of patients with an absolute, and relative decline from baseline in FVC percent predicted, and DLCO percent predicted of >10%, and >5% at months 6, 12, 18 and 24.
- Absolute change from baseline in CT visual and DTA scores at months 12 and 24.
• Absolute change from baseline in 6-minute walk distance at months 6, 12 and 24.
• Absolute change from baseline in UCSD and SGRQ score at months 6, 12, 18 and 24.
• Time to first acute FHP exacerbation, death, non-elective hospitalization or lung transplantation.
• Interaction of clinical measures of progression such as DTA scores and FVC over time.

**Data Management and Biorepository:** Data are entered at the clinical sites using an encrypted, password-protected, web-based data capture system. The data system includes internal quality checks, such as automatic range checks, to identify inconsistent, incomplete, or inaccurate data. Clinical data are entered directly from the source documents. The patient clinical data contain a unique identifying number in order to match subjects over the repeated measures. De-identified CT images collected by each site are transferred electronically to the National Jewish Health Quantitative Imaging Laboratory secured database for interpretation, archiving, and quantitative analysis of the images. Biospecimens are processed and labeled with barcodes that include the subjects electronically generated study code and the date of sample collection. The biospecimens are stored in the National Jewish Health Clinical Translational Research Center.

**Analysis Plan**

**Discovery cohort**
**Differential expression and predictive modeling:** To identify transcriptional signatures at baseline and over time in a Discovery Cohort of 85 individuals, total RNA will be examined for quality and processed for RNASeq next-generation library construction. Once constructed, RNA libraries will be sequenced on the Illumina HiSeq2500. The sequencing data will be mapped to the human reference genome and computational analysis will be performed to identify transcripts associated with progressive vs. non-progressive FHP.

Our primary objective is to determine a parsimonious transcriptional signature at presentation that is predictive of disease progression within two years of presentation. We achieve this through a two-step process: create a list of candidate genes and prune the list via feature selection. The first step uses generalized linear models to test for differential expression (DE) at initial presentation and over time between subjects with and without disease progression, controlling for additional confounding covariates (e.g., age, antigen exposure status, therapy). Here, the continuous gene expression value is dependent on progressor vs. non-progressor status to increase power. The set of DE genes identified here serves two purposes where the first use is in pathway analysis and biological function characterization to highlight biological processes that might be underlying disease progression. Secondly, the normalized expression values for the DE genes will be used as features in models predicting progression status now as the dependent variable and will undergo feature selection to find a pruned set of predictive biomarkers. We will create an ensemble model using two complementary predictive frameworks: elastic net logistic regression and random forest classifiers. Models with baseline measurements of clinical parameters, such as FVC % predicted and CT extent
of fibrosis will be compared to models with gene expression alone or in combination with clinical variables to determine any potential improvement using expression data. Prediction performance will be measured by K-fold cross-validation with the F1-measure and receiver operating characteristic curve (ROC) metrics. The candidate biomarker list will be pruned within the elastic-net framework as those features with non-zero coefficients, while features in the random forest can be ranked by their variable importance measure and recursively eliminated. The most parsimonious subset of genes that attains optimal predictive accuracy will be considered the final candidate transcriptional signature based on baseline expression for the elastic net or random forest independently. A final predictive model will be created as an ensemble of these two classifiers.

For the study’s secondary objective, a multi-step process will be conducted to identify a transcriptional profile predictive of progression that now also incorporates longitudinal changes in gene expression. Using generalized linear mixed effects models, gene expression at presentation, 12 months, and 24 months for each subject will be used as the longitudinal outcome using a categorical time variable. While also adjusting for confounders, the primary covariates of interest will be progression status-by-time interaction terms. The time trends will be split into linear and quadratic components, and the genes will be tested for differences in changes in expression patterns over time between progressors and non-progressors. This set of genes will not only seed the predictive models to be fit in the next step but also form the basis for pathway analysis, as outlined for the primary objective. In this next step, we will again fit predictive models for disease progression but now will use both linear and/or quadratic change over time.
in expression as features rather than just baseline expression. The feature selection mechanisms of the elastic net and random forests will again be used to prune the candidate set further. The performance of models using transcriptional profile information will again be compared using F1 and ROC metrics, and the most parsimonious set of gene features that achieve optimal performance will be retained for elastic nets and random forests, and an ensemble of the two will form the final prediction. Signature genes can then be distinguished by whether they contribute baseline, linear, or quadratic features to the final models. Additionally, with the repeated RNASeq measurements, we will be able to investigate how changes in gene expression over time are related to changes to specific clinical characteristics over time (e.g., CT changes over time).

**Unsupervised clustering for molecular phenotypes:** As a secondary analysis, the transcriptomic signature data will be hierarchically clustered using Ward's linkage on correlation to identify subclusters of subjects. The molecular phenotypes identified in this manner will then be used as groups and subjected to the same analysis used above with statistical and machine learning methods used above to distinguish progressors from non-progressors, this time with the intent to identify the components of the gene signature responsible for the molecular subphenotypes. These molecular phenotypes will also be associated with clinical measures of disease progression, using linear mixed models, logistic regression or a Cox proportional hazard model.

**Power:** Assuming a total sample size of 85 subjects with approximately 30% showing progression, we will have 80% power at the 0.0001 level to detect a log fold change
of 0.63 assuming an expected count of 10 (i.e., sequencing depth) and coefficient of variation of 0.43 (Figure 2).

Validation Cohort

To validate the transcriptomic signatures, we will assay gene expression via a targeted NGS panel of genes (AmpliSeq) identified with predictive differential transcriptional signatures. For this purpose, a target number of 200 genes will be selected for the AmpliSeq panel among genes prioritized using the following criteria: 1) the top genes by whole transcriptome analysis in aim 1a between progressors and non-progressors according to p-value, fold change (effect size), or base mean expression; 2) an enrichment of genes among the top genes also annotated to biological pathways and functions important for lung biology; 3) an equal number of up and down regulated genes for progressors vs. non-progressors at initial presentation; and 4) the degree to which each top gene’s expression offers predictive power for predicting progression in the models developed for the primary objective. The AmpliSeq RNA gene panel will first be run on RNA samples from progressors and non-progressors from the Discovery cohort to compare the AmpliSeq results with the corresponding whole transcriptome data, thus enabling direct analysis and/or translation across sequencing technologies. The AmpliSeq panel will then be run with isolated RNA from PBMCs of subjects in the Validation cohort (n=50) at the initial presentation and over time. Total RNA will be examined for quality and processed for AmpliSeq next-generation library construction. Once constructed, RNA libraries will be sequenced and then mapped to the human reference genome and computational analysis performed similarly to as the primary objective.
**Power:** Assuming a total sample size of 50 subjects with approximately 30% showing progression, we will have 80% power at the 0.0001 level to detect a log fold change of 0.69 assuming an expected count of 100 in the AmpliSeq panel and coefficient of variation of 0.43 (Figure 2).

**DISCUSSION**

While substantial progress has been made in understanding the clinical, radiological, and pathological manifestations of FHP, it remains difficult for the clinician to predict their clinical course or their response to therapy. Thus, the overall goal of the PREDICT study is to combine transcriptomic and clinical findings in patients with FHP to develop and validate phenotypically anchored transcriptome signatures that serve to refine the risk stratification for this complex disease.

Gene expression profiling using biological samples from carefully phenotyped clinical cohorts is a valuable tool for biomarker discovery. Five retrospective studies have evaluated the value of lung tissue gene expression profiling in HP. In a landmark study, Selman and colleagues profiled lung biopsies from 15 patients with idiopathic pulmonary fibrosis (IPF), 12 with HP, and 8 with nonspecific interstitial pneumonia and found that gene expression patterns distinguish subjects with IPF and HP. Their results showed that disease-specific gene expression signatures could provide important information on the mechanisms underlying these diseases and potentially aid in their differential diagnosis. In the largest study to date (82 lung samples with HP, 103 with IPF and 103 unaffected controls) Furusawa and colleagues replicated these findings. Also, they demonstrated that expression and pathway analysis illustrated
variations in disease severity and pathological and radiological features, highlighting both the overlap and distinct biological features of HP and IPF. Horimasu and colleagues also investigated the gene expression of nine subjects with progressive FHP and pathways related to inflammatory responses were differentially expressed compared with controls. Recently, De Sadeleer and colleagues analyzed nine FHP explant lungs and six unused donor lungs as controls and identified six molecular traits that characterize the morphological progression of FHP and associate with in vivo clinical behavior. Lastly, by integrating the expression profiling of surgical lung biopsy specimens from a patient with FHP and published study databases, Wang and colleagues identified novel functional gene candidates in FHP and the contributions of specific cell subsets relevant to FHP.

Due to the critical need for prognostic risk stratification in understanding the FHP course in a minimally invasive manner, we profiled gene expression from PBMCs in 37 patients with progressive and non-progressive FHP according to change from initial presentation in functional and radiological data within 24 months. Using a cross-validation method, we demonstrate that including baseline gene expression signature data leads to a significant increase in the prediction accuracy and AUC compared with that by clinical parameters alone (baseline FVC%, DLCO% and chest CT fibrosis) or compared with existing signatures of IPF. Hierarchical clustering applied to 74 DE transcripts shows distinct subgroups among subjects, distinguishing patients with disease progression from patients with more stable disease regardless of baseline disease severity. Collectively, these data provide a key part of the foundation for this study.
We recognize that there might not be a complete overlap in gene expression signatures between blood and lung tissue in HP. Therefore, we will determine whether the most strongly weighted genes from the transcriptomic PBMC signatures identified and validated in this study are expressed in bronchoalveolar lavage from a subset of FHP patients with disease progression. Since the primary goal is to better predict FHP disease course by discovering potential biomarkers, examining PBMC, rather than a risk-associated and difficult-to-obtain tissue such as bronchoalveolar lavage, is a priority and required for long-term viability as a clinical test. The PREDICT prospective design is the only way of establishing the incidence of disease progression (i.e., absolute risk), which is the primary study outcome to be predicted. This design also prevents measurements from being biased by knowledge of the outcome of interest. Lastly, while the integration of longitudinal analysis with whole-transcriptome profiling by using an Illumina platform-based NGS methodology (RNASeq) will be employed to identify novel, multiparametric molecular signatures that can more accurately define the broad spectrum of phenotypes characterizing the heterogeneity of FHP disease course and outcomes, the PREDICT validation study will utilize an alternative, NGS-based, custom amplicon-targeted approach with the Ion platform (AmpliSeq), thus, enabling platform-agnostic testing of potential multiparametric FHP transcriptome signatures while simultaneously providing the basis for future testing of targeted diagnostic RNA panels. In conclusion, by probing the PBMC transcriptome at baseline and during the course of the disease, the PREDICT study may ultimately result in the development of a composite index, including clinical parameters added to molecular information derived from peripheral blood which could help in patient stratification, clinical decision-making,
and lay the groundwork to define molecular targets relevant to FHP disease progression. Once established, these HP molecular signatures could be used to enrich clinical trials. Moreover, this study will generate large amounts of clinical data and blood samples that could be of great value to better understanding the role of gene-environment interactions and support integrated studies using omics data to develop personalized approaches to FHP.

ACKNOWLEDGMENTS

The PREDICT study investigators thank Beth Canono, the interstitial lung disease clinic and the adult clinical research unit at National Jewish Health for the support and the patients who participated in the study.

The PREDICT Study Investigators and Collaborators:
Evans Fernández Pérez, Sonia Leach, Brian Vestal, Lisa Maier, Kaitlin Fier, Elizabeth Del Real, Adrea Arthrofer, David Lynch, Steve Groshong, National Jewish Health; Teng Moua, and Shannon Daley, Mayo Clinic, MN; Namita Sood, Elena Foster and Anastasia Sergill, University of California, Davis; Traci Adams, Rhoda Gordon and Brian Morfin, University of Texas Southwestern Medical Center; Sachin Chaudhary and Heidi Erickson, University of Arizona College of Medicine; Mary Beth Scholand, and Chloe Kirkpatrick, University of Utah; Ayodeji Adegunsoye, Spring Maleckar, and Karina Mak, University of Chicago.
REFERENCES


Table 1. Inclusion and exclusion criteria.

<table>
<thead>
<tr>
<th><strong>Inclusion criteria</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Aged between 18 and 80 years.</td>
<td></td>
</tr>
<tr>
<td>2. Evidence of lung fibrosis on high-resolution CT performed during screening or within 4 months of enrollment and confirmed by a blinded thoracic radiologist:</td>
<td></td>
</tr>
<tr>
<td>• Reticular abnormalities and/or</td>
<td></td>
</tr>
<tr>
<td>• Traction bronchiectasis and/or</td>
<td></td>
</tr>
<tr>
<td>• Architectural distortion and/or honeycombing</td>
<td></td>
</tr>
<tr>
<td>• Lack of features suggesting an alternative diagnosis.</td>
<td></td>
</tr>
<tr>
<td>3. Have a diagnosis of HP or a provisional high-confidence diagnosis established by consensus criteria or confirmed by histopathology in the absence of other causes.</td>
<td></td>
</tr>
<tr>
<td>4. Able to understand and sign a written informed consent form.</td>
<td></td>
</tr>
<tr>
<td>5. Able to understand the importance of adherence to the study protocol and willing to follow all study requirements and undergo study procedures.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Exclusion Criteria</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Not a suitable candidate for enrollment or unlikely to comply with the requirements of this study (e.g., unstable or deteriorating cardiac disease, dementia).</td>
<td></td>
</tr>
<tr>
<td>2. Known explanation for the interstitial lung disease, including but not limited to radiation, drug toxicity, sarcoidosis, pneumoconiosis.</td>
<td></td>
</tr>
<tr>
<td>3. Concomitant clinical diagnosis of a connective tissue disease, including but not limited to scleroderma, polymyositis/dermatomyositis and rheumatoid arthritis.</td>
<td></td>
</tr>
<tr>
<td>4. Listed or expected to receive a lung transplant within 4-6 months from enrollment.</td>
<td></td>
</tr>
<tr>
<td>5. Active smoker.</td>
<td></td>
</tr>
<tr>
<td>6. Lymphoproliferative disease or malignancies other than nonmetastatic skin cancer.</td>
<td></td>
</tr>
<tr>
<td>7. Pregnant women.</td>
<td></td>
</tr>
</tbody>
</table>
**Table 2. Schedule of activities.**

<table>
<thead>
<tr>
<th>Data and Sample Collection</th>
<th>Screening</th>
<th>Visit 1 Month 0</th>
<th>Visit 2 6 months</th>
<th>Visit 3 12 months</th>
<th>Visit 4 18 months</th>
<th>Visit 5 24 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informed consent</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demographics</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Medical history and physical exam</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Environmental and occupational questionnaire</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>History, vital signs and physical exam</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Medication review, hospitalization and respiratory exacerbation assessments, vital status</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Patient reported outcome measure questionnaires</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Lung volumes, spirometry and DLCO</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Six-minute walk test</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Chest high-resolution CT*</td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Baseline histopathology**</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blood sample collection</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Complete case report forms</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

*All HRCT reviewed and recorded by a blinded central reader.

**All diagnostic histopathology reviewed and recorded by a blinded central reader.
Figure 1. Study design.

Total N = 135
Pre-screen potential participants by inclusion and exclusion criteria

Conduct informed consent. Perform baseline assessments then followed over 24 months. Refer to Table 1, Schedule of Activities.

Discovery N = 85
Illumina Whole Transcriptome FNASeq FBMC signature at initial presentation and at month 12 and 24 in progressors and non-progressors

Validation N = 50
Illumina Whole Transcriptome FNASeq FBMC signature at initial presentation and at month 12 and 24 in progressors and non-progressors

Clinical Transcriptomic Signature for FHP Risk Stratification
Figure 2. Study power.