



(51) International Patent Classification:

C12P 19/34 (2006.01) G01N 33/569 (2006.01)

C12Q 1/68 (2018.01) G01N 33/68 (2006.01)

G01N 33/53 (2006.01)

(21) International Application Number:

PCT/US2020/013367

(22) International Filing Date:

13 January 2020 (13.01.2020)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/791,083 11 January 2019 (11.01.2019) US

62/849,630 17 May 2019 (17.05.2019) US

(71) Applicant: UNIVERSITY OF VIRGINIA PATENT FOUNDATION [US/US]; 722 Preston Avenue, Suite 107, Charlottesville, VA 22903 (US).

(72) Inventor: NOTH, Imre; University Of Virginia Hospital Dept. Of Medicine, /pulmonary & Critical Care, 1 Hospital Drive, Clinical Dept. Wing, 5th Floor, Rm 5592 Po Box 800546, Charlottesville, VA 22908 (US).

(74) Agent: PERKINS, Christopher, P.; Jenkins, Wilson, Taylor & Hunt, P.A., 3015 Carrington Mill Boulevard, Suite 550, Morrisville, NC 27560 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP,

(54) Title: COMPOSITIONS AND METHODS FOR PREDICTING LUNG FUNCTION DECLINE IN IDIOPATHIC PULMONARY FIBROSIS

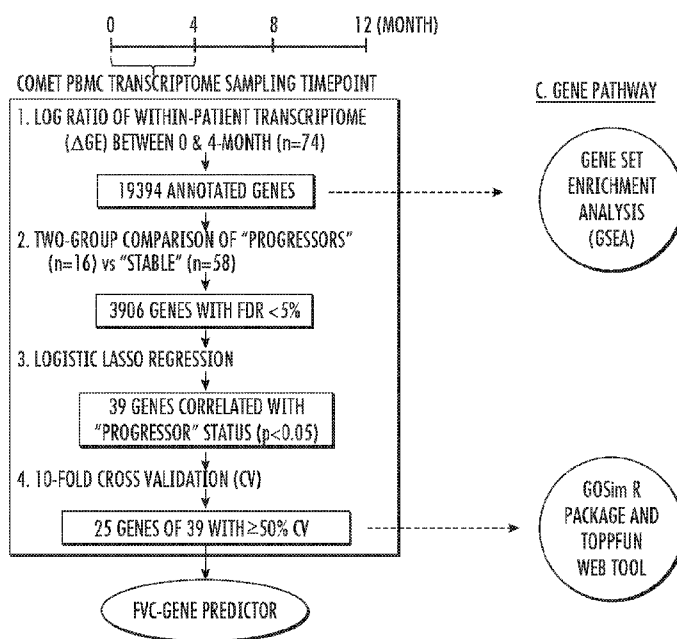


FIG. 1A

(57) Abstract: Provided are methods for generating prognostic signatures for subject diagnosed with Idiopathic Pulmonary Fibrosis (IPF) with respect to decline in lung Forced Vital Capacity (FVC). The methods can include determining first expression levels for one or more genes as set forth herein in a first biological sample obtained from a subject diagnosed with IPF, determining a second expression level for the same one or more genes in a second biological sample obtained from the subject, and comparing the first and second expression levels for the one or more genes to provide a prognostic signature. The first and second biological samples can include peripheral blood mononuclear cells (PBMCs) and/or nucleic acids extracted from PBMCs. Also provided are methods for classifying subjects with IPF as being at risk for FVC decline, for identifying and treating at risk subjects, and for monitoring the progress of treatments.

KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

- (84) Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

DESCRIPTION

COMPOSITIONS AND METHODS FOR PREDICTING LUNG FUNCTION DECLINE IN IDIOPATHIC PULMONARY FIBROSIS

CROSS REFERENCE TO RELATED APPLICATION

The presently disclosed subject matter claims the benefit of U.S. Provisional Patent Application Serial Nos. 62/791,083, filed January 11, 2019; and 62/849,630, filed May 17, 2019; the disclosure of each of which is incorporated herein by reference in its entirety.

GOVERNMENT INTEREST

This invention was made with government support under grant number HL130796 awarded by The National Institutes of Health. The government has certain rights in the invention.

BACKGROUND

Idiopathic pulmonary fibrosis (IPF) is a deadly and progressive scarring lung disease of unknown etiology (Raghu et al., 2011). Prior to death, most patients experience progressive lung function decline, as measured by forced vital capacity (FVC). Longitudinal decline in FVC is a well-validated predictor of mortality and is often used as the primary efficacy endpoint in IPF clinical trials (du Bois et al., 2011; Schmidt et al., 2014; Karimi-Shah & Chowdhury, 2015). However, while most patients experience FVC decline, the rate is variable and periods of relative FVC stability are also often observed. Such heterogeneity hampers the development of effective therapies (Kaner et al., 2019), as many patients do not experience FVC decline during the clinical trial period (King et al., 2014; Richeldi et al., 2014).

Efforts to identify patients who will experience FVC decline have been met with frustration. Clinical prediction models reliably predict increased mortality risk, but fail to accurately predict FVC decline (Ley et al., 2016). Several genetic and plasma biomarkers have also been linked with mortality (Greene et al., 2002; Rosas et al., 2008; Richards et al., 2012; Herazo-Maya et al., 2013; Peljto et al., 2013; Ley et al., 2014; Herazo-Maya et al., 2017), but have less association with FVC decline. Even FVC decline itself fails to predict future FVC decline (Jegal et al., 2005; Schmidt et al., 2014). Such observations suggest that FVC reflects established fibrotic remodeling rather than ongoing, potentially modifiable processes leading to fibrosis. As such, markers of disease activity, rather than severity, would be of use in informing diagnoses.

The dynamic nature of the transcriptome has the potential to signal early indications of fibrosis activity. The great potential of this approach was previously demonstrated with a 52-gene signature that predicted IPF survival in cohorts around the world (Herazo-Maya et al., 2013; Herazo-Maya et al., 2017). This gene signature relied on cross-sectional data, which like clinical prediction models, may not account for critical gene expression changes that likely occur with disease activity. In accordance with the presently disclosed subject matter, it was investigated whether longitudinal within-patient gene expression changes would reflect disease activity, as measured by FVC decline. Using the presently disclosed results, a transcriptomic predictor of FVC decline was developed and validated in three independent IPF cohorts.

SUMMARY

This Summary lists several embodiments of the presently disclosed subject matter, and in many cases lists variations and permutations of these embodiments of the presently disclosed subject matter. This Summary is merely exemplary of the numerous and varied embodiments. Mention of one or more representative features of a given embodiment is likewise exemplary. Such an embodiment can typically exist with or without the feature(s) mentioned; likewise, those features can be applied to other embodiments of the presently disclosed subject matter, whether listed in this Summary or not. To avoid excessive repetition, this Summary does not list or suggest all possible combinations of such features.

In some embodiments, the presently disclosed subject matter pertains to methods for generating prognostic signatures for subjects diagnosed with Idiopathic Pulmonary Fibrosis (IPF) with respect to decline in lung Forced Vital Capacity (FVC). In some embodiments, the presently disclosed methods comprise determining a first expression level for one or more genes selected from the group consisting of *ALDH4A1*, *APTX*, *ATP6AP1L*, *CCNB1*, *CNR2*, *DNAJC17*, *DTWD1*, *FAM111B*, *GABRR1*, *GPR39*, *GYP A*, *HBB*, *HLA-DPB1*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MRPL35*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PLCL1*, *PNMA5*, *RAB3C*, *RBM43*, *RLBP1*, *SESN3*, *SLC25A37*, *SSU72P8*, *TP63*, *WDR17*, *ZNF252P*, and *ZNF582* in a first biological sample obtained from the subject diagnosed with IPF to establish a baseline expression level for the one or more genes; determining a second expression level for the one or more genes in a second biological sample obtained from the subject, wherein the first and second biological samples comprise peripheral blood mononuclear cells (PBMCs) and/or nucleic acids extracted from PBMCs; and comparing the first and second expression levels for the one or more genes, wherein the comparing

provides a prognostic signature for the subject with respect to decline in lung FVC within two years from the time that the first biological sample was obtained from the subject. In some embodiments, the presently disclosed methods comprise determining first and second expression levels for a set of genes selected from the group consisting of (a) *APTX*, *CNR2*, *GYP A*, *ITLN1*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PLA2G4A*, and *PNMA5*; (b) *APTX*, *ATP6AP1L*, *ITLN1*, *LINC00319*, *MAZ*, *MSR1*, *NT5E*, *PCDHB15*, *RAB3C*, *SSU72P8*, and *TP62*; (c) *APTX*, *CNR2*, *GABRR1*, *GPR39*, *GYP A*, *HBB*, *ITLN1*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PNMA5*, *RLBP1*, and *SSU72P8*; and (d) *APTX*, *ATP6AP1L*, *CNR2*, *FAM111B*, *GABRR1*, *GPR39*, *GYP A*, *HBB*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PLCL1*, *PNMA5*, *RAB3C*, *RBM43*, *RLBP1*, *SSU72P8*, *TP63*, and *ZNF252P*. In some embodiments, the presently disclosed methods comprise determining first and second expression levels for each of *APTX*, *ATP6AP1L*, *CNR2*, *FAM111B*, *GABRR1*, *GPR39*, *GYP A*, *HBB*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PLCL1*, *PNMA5*, *RAB3C*, *RBM43*, *RLBP1*, *SSU72P8*, *TP63*, and *ZNF252P*.

In some embodiments, the second biological sample is obtained from the subject at a time from about 4 to about 12 months subsequent to when the first biological sample was obtained from the subject.

In some embodiments, the subject is a human.

In some embodiments, one or both determining steps comprise a technique selected from the group consisting of RNA-seq analysis, quantitative polymerase chain reaction (PCR) including quantitative reverse transcription PCR (qRT-PCR), and the use of a nucleic acid or protein array, or any combination thereof.

In some embodiments, the comparing step comprises comparing a normalized expression level for each gene in the first biological sample to a normalized expression level for each gene in the second biological sample to generate a fold-increase and/or a fold-decrease in the second biological sample relative to the first biological sample for each gene. In some embodiments, the comparing step comprises summing each fold-increase and/or fold-decrease to produce an FVC-gene predictor score for the subject.

In some embodiments, the summing is performed after multiplying each fold-increase and/or fold-decrease by a weighting value to produce a weighted FVC-gene predictor score for the subject.

The presently disclosed subject matter also related in some embodiments to methods for classifying subjects diagnosed with IPF as being at risk for FVC decline. In some embodiments, the methods comprise determining a first expression level for one or more genes selected from the group consisting of *ALDH4A1*, *APTX*, *ATP6AP1L*, *CCNB1*, *CNR2*, *DNAJC17*, *DTWD1*, *FAM111B*, *GABRR1*, *GPR39*, *GYP A*, *HBB*, *HLA-DPB1*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MRPL35*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PLCL1*, *PNMA5*, *RAB3C*, *RBM43*, *RLBP1*, *SESN3*, *SLC25A37*, *SSU72P8*, *TP63*, *WDR17*, *ZNF252P*, and *ZNF582* in a first biological sample obtained from the subject diagnosed with IPF to establish a baseline expression level for the one or more genes; determining a second expression level for the one or more genes in a second biological sample obtained from the subject, wherein the first and second biological samples comprise peripheral blood mononuclear cells (PBMCs) and/or nucleic acids extracted from PBMCs; and comparing the first and second expression levels for the one or more genes to create an FVC-gene predictor score, wherein if the FVC-gene predictor score is greater than or equal to a pre-selected value, the patient is classified as being at risk for a decline in lung FVC within two years from the time that the first biological sample was obtained from the subject. In some embodiments, the comparing comprises comparing a normalized expression level for each gene in the first biological sample to a normalized expression level for each gene in the second biological sample to generate a fold-increase and/or a fold-decrease in the second biological sample relative to the first biological sample for each gene. In some embodiments, the comparing further comprises summing each fold-increase and/or fold-decrease to produce an FVC-gene predictor score for the subject. In some embodiments, the summing is performed after multiplying each fold-increase and/or fold-decrease by a weighting value to produce a weighted FVC-gene predictor score for the subject. In some embodiments, the presently disclosed subject matter methods comprise determining first and second expression levels for a set of genes selected from the group consisting of (a) *APTX*, *CNR2*, *GYP A*, *ITLN1*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PLA2G4A*, and *PNMA5*; (b) *APTX*, *ATP6AP1L*, *ITLN1*, *LINC00319*, *MAZ*, *MSR1*, *NT5E*, *PCDHB15*, *RAB3C*, *SSU72P8*, and *TP63*; (c) *APTX*, *CNR2*, *GABRR1*, *GPR39*, *GYP A*, *HBB*, *ITLN1*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PNMA5*, *RLBP1*, and *SSU72P8*; and (d) *APTX*, *ATP6AP1L*, *CNR2*, *FAM111B*, *GABRR1*, *GPR39*, *GYP A*, *HBB*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PLCL1*, *PNMA5*, *RAB3C*, *RBM43*, *RLBP1*, *SSU72P8*, *TP63*, and *ZNF252P*. In some embodiments, the presently disclosed subject

matter methods comprise determining first and second expression levels for each of *APTX*, *ATP6AP1L*, *CNR2*, *FAM111B*, *GABRR1*, *GPR39*, *GYPA*, *HBB*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PLCL1*, *PNMA5*, *RAB3C*, *RBM43*, *RLBP1*, *SSU72P8*, *TP63*, and *ZNF252P*.

In some embodiments, the second biological sample is obtained from the subject at a time from about 4 to about 12 months subsequent to when the first biological sample was obtained from the subject.

In some embodiments, the subject is a human.

In some embodiments, one or both determining steps comprises a technique selected from the group consisting of RNA-seq analysis, quantitative polymerase chain reaction (PCR) including quantitative reverse transcription PCR (qRT-PCR), and the use of a nucleic acid or protein array, or any combination thereof.

The presently disclosed subject matter also relates in some embodiments to methods for identifying and treating subjects diagnosed with IPF and/or who are at risk for a decline in lung Forced Vital Capacity (FVC). In some embodiments, the methods comprise determining a first expression level for one or more genes selected from the group consisting of *ALDH4A1*, *APTX*, *ATP6AP1L*, *CCNB1*, *CNR2*, *DNAJC17*, *DTWD1*, *FAM111B*, *GABRR1*, *GPR39*, *GYPA*, *HBB*, *HLA-DPB1*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MRPL35*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PLCL1*, *PNMA5*, *RAB3C*, *RBM43*, *RLBP1*, *SESN3*, *SLC25A37*, *SSU72P8*, *TP63*, *WDR17*, *ZNF252P*, and *ZNF582* in a first biological sample obtained from the subject diagnosed with IPF to establish a baseline expression level for the one or more genes; determining a second expression level for the one or more genes in a second biological sample obtained from the subject, wherein the first and second biological samples comprise peripheral blood mononuclear cells (PBMCs) and/or nucleic acids extracted from PBMCs; comparing the first and second expression levels for the one or more genes to create an FVC-gene predictor score; and if the FVC-gene predictor score is greater than or equal to a pre-selected value, treating the subject with a treatment selected from the group consisting of lung transplantation and a drug therapy. In some embodiments, the drug therapy comprises administering to the subject a pharmaceutical composition comprising pirfenidone, nintedanib, or a combination thereof in an amount and via a route of administration effective to delay or prevent the development of FVC decline in the subject. In some embodiments, the comparing comprises comparing a normalized expression level for each gene in the first biological sample to a normalized expression level

for each gene in the second biological sample to generate a fold-increase and/or a fold-decrease in the second biological sample relative to the first biological sample for each gene. In some embodiments, the comparing further comprises summing each fold-increase and/or fold-decrease to produce an FVC-gene predictor score for the subject. In some embodiments, the summing is performed after multiplying each fold-increase and/or fold-decrease by a weighting value to produce a weighted FVC-gene predictor score for the subject. In some embodiments, the presently disclosed methods comprise determining first and second expression levels for a set of genes selected from the group consisting of (a) *APTX*, *CNR2*, *GYP A*, *ITLN1*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PLA2G4A*, and *PNMA5*; (b) *APTX*, *ATP6AP1L*, *ITLN1*, *LINC00319*, *MAZ*, *MSR1*, *NT5E*, *PCDHB15*, *RAB3C*, *SSU72P8*, and *TP62*; (c) *APTX*, *CNR2*, *GABRR1*, *GPR39*, *GYP A*, *HBB*, *ITLN1*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PNMA5*, *RLBP1*, and *SSU72P8*; and (d) *APTX*, *ATP6AP1L*, *CNR2*, *FAM111B*, *GABRR1*, *GPR39*, *GYP A*, *HBB*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PLCL1*, *PNMA5*, *RAB3C*, *RBM43*, *RLBP1*, *SSU72P8*, *TP63*, and *ZNF252P*. In some embodiments, the presently disclosed methods comprise determining first and second expression levels for each of *APTX*, *ATP6AP1L*, *CNR2*, *FAM111B*, *GABRR1*, *GPR39*, *GYP A*, *HBB*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PLCL1*, *PNMA5*, *RAB3C*, *RBM43*, *RLBP1*, *SSU72P8*, *TP63*, and *ZNF252P*.

In some embodiments, the second biological sample is obtained from the subject at a time from about 4 to about 12 months subsequent to when the first biological sample was obtained from the subject.

In some embodiments, the subject is a human.

In some embodiments, one or both determining steps comprise a technique selected from the group consisting of RNA-seq analysis, quantitative polymerase chain reaction (PCR) including quantitative reverse transcription PCR (qRT-PCR), and the use of a nucleic acid or protein array, or any combination thereof.

In some embodiments, the presently disclosed subject matter also relates to methods for monitoring the progress of a treatment in an IPF patient whose is experiencing a decline in lung Forced Vital Capacity FVC. In some embodiments, the method comprises determining a first expression level for one or more genes selected from the group consisting of *ALDH4A1*, *APTX*, *ATP6AP1L*, *CCNB1*, *CNR2*, *DNAJC17*, *DTWD1*, *FAM111B*, *GABRR1*, *GPR39*, *GYP A*, *HBB*, *HLA-DPB1*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MRPL35*,

MSR1, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PLCL1*, *PNMA5*, *RAB3C*, *RBM43*, *RLBP1*, *SESN3*, *SLC25A37*, *SSU72P8*, *TP63*, *WDR17*, *ZNF252P*, and *ZNF582* in a first biological sample obtained from the patient to establish a baseline expression level for the one or more genes; determining a second expression level for the one or more genes in a second biological sample obtained from the patient at a subsequent time point, wherein the first and second biological samples comprise peripheral blood mononuclear cells (PBMCs) and/or nucleic acids extracted from PBMCs; and comparing the first and second expression levels for the one or more genes, wherein the comparing step is indicative of the progress of the treatment in the patient. In some embodiments, the treatment comprises administering to the patient a pharmaceutical composition comprising pirfenidone, nintedanib, or a combination. In some embodiments, the comparing comprises comparing a normalized expression level for each gene in the first biological sample to a normalized expression level for each gene in the second biological sample to generate a fold-increase and/or a fold-decrease in the second biological sample relative to the first biological sample for each gene. In some embodiments, the comparing further comprises summing each fold-increase and/or fold-decrease to produce an FVC-gene predictor score for the patient. In some embodiments, the summing is performed after multiplying each fold-increase and/or fold-decrease by a weighting value to produce a weighted FVC-gene predictor score for the patient. In some embodiments, the presently disclosed methods comprise determining first and second expression levels for a set of genes selected from the group consisting of (a) *APTX*, *CNR2*, *GYPA*, *ITLN1*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PLA2G4A*, and *PNMA5*; (b) *APTX*, *ATP6AP1L*, *ITLN1*, *LINC00319*, *MAZ*, *MSR1*, *NT5E*, *PCDHB15*, *RAB3C*, *SSU72P8*, and *TP62*; (c) *APTX*, *CNR2*, *GABRR1*, *GPR39*, *GYPA*, *HBB*, *ITLN1*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PNMA5*, *RLBP1*, and *SSU72P8*; and (d) *APTX*, *ATP6AP1L*, *CNR2*, *FAM111B*, *GABRR1*, *GPR39*, *GYPA*, *HBB*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PLCL1*, *PNMA5*, *RAB3C*, *RBM43*, *RLBP1*, *SSU72P8*, *TP63*, and *ZNF252P*. In some embodiments, the presently disclosed methods comprise determining first and second expression levels for each of *APTX*, *ATP6AP1L*, *CNR2*, *FAM111B*, *GABRR1*, *GPR39*, *GYPA*, *HBB*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PLCL1*, *PNMA5*, *RAB3C*, *RBM43*, *RLBP1*, *SSU72P8*, *TP63*, and *ZNF252P*.

In some embodiments, the second biological sample is obtained from the patient at a time subsequent to when the first biological sample was obtained from the patient selected

from the group consisting of about 1 week, about 2 weeks, about 4 weeks, about 6 weeks, about 2 months, about 3 months, about 4 months, about 5 months, about 6 months, or longer than six months.

In some embodiments, the patient is a human.

In some embodiments, one or both determining steps comprise a technique selected from the group consisting of RNA-seq analysis, quantitative polymerase chain reaction (PCR) including quantitative reverse transcription PCR (qRT-PCR), and the use of a nucleic acid or protein array, or any combination thereof.

In some embodiments, the presently disclosed methods further comprise determining a one or more subsequent expression levels for one or more genes selected from the group consisting of *ALDH4A1*, *APTX*, *ATP6AP1L*, *CCNB1*, *CNR2*, *DNAJC17*, *DTWD1*, *FAM111B*, *GABRR1*, *GPR39*, *GYPA*, *HBB*, *HLA-DPB1*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MRPL35*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PLCL1*, *PNMA5*, *RAB3C*, *RBM43*, *RLBP1*, *SESN3*, *SLC25A37*, *SSU72P8*, *TP63*, *WDR17*, *ZNF252P*, and *ZNF582* in one or more subsequently isolated biological samples obtained from the patient; and comparing the first, second, and one or more subsequent expression levels for the one or more genes, wherein the comparing step is indicative of the progress of the treatment in the patient.

Accordingly, it is an object of the presently disclosed subject matter to provide compositions and methods for predicting lung function decline in patients with idiopathic pulmonary fibrosis. This and other objects are achieved in whole or in part by the presently disclosed subject matter. Further, objects of the presently disclosed subject matter having been stated above, other objects and advantages of the presently disclosed subject matter will become apparent to those skilled in the art after a study of the following description, Figures, and EXAMPLES. Additionally, various aspects and embodiments of the presently disclosed subject matter are described in further detail below.

BRIEF DESCRIPTION OF THE FIGURES

Figures 1A and 1B. Summary of the development and validation of the FVC-gene predictor. Figure 1A. Flowchart of the COMET training cohort: The left panel summarized exemplary steps for identifying the 25-gene FVC-gene predictor predictive of FVC event status using short-term (0-4 month) within-patient Δ GE. The right panel shows gene pathways analyses applied to entire annotated gene set and to 25 genes that constituted the predictor. Figure 1B. COMET subsets and independent validation cohorts with different transcriptome assay platforms: Steps of testing varied transcriptome sampling timepoints

and durations in COMET subsets (left) and testing in external independent cohorts (right) using overlapping Gene ID in differing transcriptome assay platforms. Scores were determined as continuous values using the CV weights for each FVC predictor gene in each patient.

Figures 2A-2C. Classification of COMET training cohort using the genes constituted FVC-gene predictor. Figure 2A. Hierarchical clustering of the 74 IPF patients in COMET. Below the clustering are three sets of data, with individual patients depicted as a horizontal line in Figure 2A. The first group of vertical lines show the gender of the patient, with females in gray and males in black. The second group of vertical lines show the FVC status of each patient, with FVC stable patients in light gray (green in the color version of Figure 2A) and FVC decline patients in dark gray (blue in the color version of Figure 2A). The third group of vertical lines show the DLCO status of each patient, with DLCO stable patients in white (yellow in the color version of Figure 2A) and DLCO decline patients in light gray (light blue in the color version of Figure 2A). All of the 16 FVC progressor (FVC=1) patients (blue in the color version of Figure 2A; dark gray in the black and white version of Figure 2A) were enriched in the bottom cluster, while the upper cluster only contained FVC stable (FVC=0) patients (green in the color version of Figure 2A; light gray in the black and white version of Figure 2A). Red, white, and blue color in the color version of Figure 2A indicate Δ GE values above, at, and below, respectively, the average Δ GE of the corresponding gene. Figures 2B and 2C. PCA of COMET training cohort based on the FVC-gene predictor using R/CRAN package “FactoMineR”. Individual factor map with confidence ellipses around FVC progressor (blue in color version of Figure 2B; black in black and white version of Figure 2B) or stable (green in color version of Figure 2B; gray in black and white version of Figure 2B) status is shown in Figure 2B. Variables factor map with predictor genes is shown in Figure 2C. The projection of the arrow head of each variable (i.e., gene) onto each dimension represents the component loadings of the corresponding gene.

Figures 3A-3C. Receiver-Operating-Characteristic (ROC) and Area Under the Curve (AUC) analysis of FVC-gene predictor. AUC values with 95% confidence intervals are displayed in right bottom of each graph. Dotted red line denotes specificity at 75%. Figure 3A. Independent validation cohorts. At anchored specificity of -75%, the sensitivities are 75.0%, 66.7% and 80.0%, for UChicago, UPMA and Imperial cohorts, respectively. Figure 3B. Training and subset the COMET cohort (I) with increasing

transcriptome sampling durations for determination of Δ GE. At anchored specificity of -75%, sensitivities are 100%, 92.3% and 78.6%, for 0-4 month, 0-8 month, and 0-12 month, respectively. FVC-gene predictor performance modestly diminished moving from 4 month to 8 months and 12 months. Figure 3C. Training and subsets of the COMET cohort (II) with 4 month transcriptome sampling but varying baselines for Δ GE determination. At anchored specificity of -75%, sensitivities are 100%, 69.2% and 36.4% for 0-4 month, 4-8 month, and 8-12 month, respectively. Performance diminishes more dramatically and use of months 8-12 are ineffective. Detailed ROC/AUC analyses results can be found in Tables 5 and 6.

Figures 4A-4C. Receiver-Operating-Characteristic (ROC) analysis and area under curve (AUC) of FVC-gene predictor in three independent validation cohorts when FVC decline event was defined as $\geq 5\%$ relative decline in FVC% of predicted. Figure 4A. University of Chicago (UChicago); Figure 4B. University of Pittsburg Medical Center (UPMC); Figure 4C. Imperial College London (Imperial). 1-specificity and sensitivity are displayed on upper left, total number of patients (FVC stable/FVC progressive) and AUC with 95% confidence intervals in parenthesis are displayed on lower right of each graph.

Figures 5A-5C. Hierarchical clustering of subsets of COMET cohort based on FVC-gene predictor. Hierarchical clustering of the IPF patients in three COMET subsets with GE sampling duration (in months) of 0-8 month (Figure 5A), 4-8 month (Figure 5B), and 0-12 month (Figure 5C), respectively. Red, white, and blue color in the color version of Figure 5A indicate Δ GE values above, at, or below the average Δ GE of the corresponding gene. 85%-92% of FVC decline (FVC=1) were enriched in the left cluster, while the FVC stable (FVC=0) patients were classified mainly in the right cluster. The FVC event duration (in months) is defined at $<$ or $\geq 10\%$ relative decline of FVC % of prediction over 12 months for all three groups.

Figures 6A-6D. ROC analysis and AUC curve of early (0-4 month) FVC change and peripheral plasma biomarkers in prediction of FVC decline at 12 month. Figure 6A. baseline FVC predicted percentage; Figure 6B. Baseline peripheral plasma MM7 (Matrix metalloproteinase-7); Figure 6C. Baseline peripheral plasma POSTN (Periostin); Figure 6D. Baseline peripheral plasma CCL18 (C-C motif chemokine ligand 18).

Figures 7A and 7B. Gene Set Enrichment Analysis (GSEA) of COMET longitudinal 0-4 month gene expression changes between progressive and stable IPF patients. Enrichment plots of significant functional gene sets expression changes of 19394

annotated genes from baseline to 4 month were analyzed for their differential functional profiles between progressive FVC and stable FVC patients. Figure 7A. Enrichment plot of 27 hallmark genes in TGF-beta signaling pathway. These genes demonstrated significant larger magnitude of changes from baseline to 4 month in progressive FVC patients than those remaining stable. Figure 7B. Enrichment plot of 10 genes in Glycan degradation activity demonstrating larger magnitude of changes from baseline to 4 month in progressive FVC patients than stable FVC patients. Core Enrichment genes in each pathway are displayed in Tables 7 and 8.

Figures 8A-8D. Coefficient of Variation (CoV) analysis and power estimation of COMET cohort data. Training data represents gene expression difference (Δ GE) between baseline and 4 month follow-up, whereas baseline data represents cross-sectional gene expressions (GE). Figures 8A and 8B. Minus versus Average plot between the CoV of baseline GE (CoV1) and the training Δ GE (CoV2) in FVC stable (Figure 8A) and FVC progressor group (Figure 8B). Figure 8C. Power estimation based on postulated sample sizes of PBMC transcriptome using R/CRAN package “sizepower”. Horizontal dotted line indicated power of 0.9, at an alpha of 0.05 whereas the corresponding sample size is 63 for baseline GE, and 16 for Δ GE. Figure 8D. Intra-subject CoV analysis across different PBMC sampling time in COMET. Black bar: Intra-Subject CoV is larger in progressor than in stable patients in COMET Δ GE data; Grey bar: Intra-Subject CoV is larger in stable patients than in progressors in COMET Δ GE data.

DETAILED DESCRIPTION

Headings are included herein for reference and to aid in locating certain sections. These headings are not intended to limit the scope of the concepts described therein under, and these concepts may have applicability in other sections throughout the entire specification.

I. Definitions

The terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting of the presently disclosed subject matter.

While the following terms are believed to be well understood by one of ordinary skill in the art, the following definitions are set forth to facilitate explanation of the presently disclosed subject matter.

All technical and scientific terms used herein, unless otherwise defined below, are intended to have the same meaning as commonly understood by one of ordinary skill in the

art. References to techniques employed herein are intended to refer to the techniques as commonly understood in the art, including variations on those techniques or substitutions of equivalent techniques that would be apparent to one of skill in the art. While the following terms are believed to be well understood by one of ordinary skill in the art, the following definitions are set forth to facilitate explanation of the presently disclosed subject matter.

The articles “a” and “an” are used herein to refer to one or to more than one (i.e., to at least one) of the grammatical object of the article. By way of example, “an element” means one element or more than one element.

The term “about”, as used herein, means approximately, in the region of, roughly, or around. When the term “about” is used in conjunction with a numerical range, it modifies that range by extending the boundaries above and below the numerical values set forth. For example, In some embodiments, the term “about” is used herein to modify a numerical value above and below the stated value by a variance of 10%. Therefore, about 50% means in the range of 45%-55%. Numerical ranges recited herein by endpoints include all numbers and fractions subsumed within that range (e.g., 1 to 5 includes 1, 1.5, 2, 2.75, 3, 3.90, 4, and 5). It is also to be understood that all numbers and fractions thereof are presumed to be modified by the term “about”.

The terms “additional therapeutically active compound” or “additional therapeutic agent”, as used in the context of the presently disclosed subject matter, refers to the use or administration of a compound for an additional therapeutic use for a particular injury, disease, or disorder being treated. Such a compound, for example, could include one being used to treat an unrelated disease or disorder, or a disease or disorder which may not be responsive to the primary treatment for the injury, disease or disorder being treated. Disease and disorders being treated by the additional therapeutically active agent include, for example, hypertension and diabetes. The additional compounds may also be used to treat symptoms associated with the injury, disease, or disorder, including, but not limited to, pain and inflammation.

The term “adult” as used herein, is meant to refer to any non-embryonic or non-juvenile subject. For example, the term “adult adipose tissue stem cell”, refers to an adipose stem cell, other than that obtained from an embryo or juvenile subject.

As used herein, an “agonist” is a composition of matter which, when administered to a mammal such as a human, enhances or extends a biological activity attributable to the level or presence of a target compound or molecule of interest in the subject.

A disease or disorder is “alleviated” if the severity of a symptom of the disease, condition, or disorder, or the frequency with which such a symptom is experienced by a subject, or both, are reduced.

As used herein, an “analog” of a chemical compound is a compound that, by way of example, resembles another in structure but is not necessarily an isomer (e.g., 5-fluorouracil is an analog of thymine).

An “antagonist” is a composition of matter which when administered to a mammal such as a human, inhibits a biological activity attributable to the level or presence of a compound or molecule of interest in the subject.

The term “antibody”, as used herein, refers to an immunoglobulin molecule which is able to specifically bind to a specific epitope on an antigen. Antibodies can be intact immunoglobulins derived from natural sources or from recombinant sources and can be immunoreactive portions of intact immunoglobulins. Antibodies are typically tetramers of immunoglobulin molecules. The antibodies in the presently disclosed subject matter may exist in a variety of forms including, for example, polyclonal antibodies, monoclonal antibodies, Fv, Fab and F(ab)₂, as well as single chain antibodies and humanized antibodies.

The term “autologous”, as used herein, refers to something that occurs naturally and normally in a certain type of tissue or in a specific structure of the body. In transplantation, it refers to a graft in which the donor and recipient areas are in the same individual, or to blood that the donor has previously donated and then receives back, usually during surgery.

The term “biological sample”, as used herein, refers to samples obtained from a living organism, including skin, hair, tissue, blood, plasma, cells, sweat, and urine.

The term “bioresorbable”, as used herein, refers to the ability of a material to be resorbed in vivo. “Full” resorption means that no significant extracellular fragments remain. The resorption process involves elimination of the original implant materials through the action of body fluids, enzymes, or cells. Resorbed calcium carbonate may, for example, be redeposited as bone mineral, or by being otherwise re-utilized within the body, or excreted. “Strongly bioresorbable”, as the term is used herein, means that at least 80% of the total mass of material implanted is resorbed within one year.

The term “clearance”, as used herein refers to the physiological process of removing a compound or molecule, such as by diffusion, exfoliation, removal via the bloodstream, and excretion in urine, or via sweat or other fluid.

A “control” cell, tissue, sample, or subject is a cell, tissue, sample, or subject of the same type as a test cell, tissue, sample, or subject. The control may, for example, be examined at precisely or nearly the same time the test cell, tissue, sample, or subject is examined. The control may also, for example, be examined at a time distant from the time at which the test cell, tissue, sample, or subject is examined, and the results of the examination of the control may be recorded so that the recorded results may be compared with results obtained by examination of a test cell, tissue, sample, or subject. The control may also be obtained from another source or similar source other than the test group or a test subject, where the test sample is obtained from a subject suspected of having a disease or disorder for which the test is being performed.

A “test” cell, tissue, sample, or subject is one being examined or treated.

A “pathoindicative” cell, tissue, or sample is one which, when present, is an indication that the animal in which the cell, tissue, or sample is located (or from which the tissue was obtained) is afflicted with a disease or disorder. By way of example, the presence of one or more breast cells in a lung tissue of an animal is an indication that the animal is afflicted with metastatic breast cancer.

A tissue “normally comprises” a cell if one or more of the cell are present in the tissue in an animal not afflicted with a disease or disorder.

A “compound”, as used herein, refers to any type of substance or agent that is commonly considered a drug, or a candidate for use as a drug, combinations, and mixtures of the above, as well as polypeptides and antibodies of the presently disclosed subject matter.

The use of the word “detect” and its grammatical variants is meant to refer to measurement of the species without quantification, whereas use of the word “determine” or “measure” with their grammatical variants are meant to refer to measurement of the species with quantification. The terms “detect” and “identify” are used interchangeably herein.

As used herein, a “detectable marker” or a “reporter molecule” is an atom or a molecule that permits the specific detection of a compound comprising the marker in the presence of similar compounds without a marker. Detectable markers or reporter molecules include, e.g., radioactive isotopes, antigenic determinants, enzymes, nucleic acids available for hybridization, chromophores, fluorophores, chemiluminescent molecules, electrochemically detectable molecules, and molecules that provide for altered fluorescence-polarization or altered light-scattering.

A “disease” is a state of health of an animal wherein the animal cannot maintain homeostasis, and wherein if the disease is not ameliorated then the animal’s health continues to deteriorate.

In contrast, a “disorder” in an animal is a state of health in which the animal is able to maintain homeostasis, but in which the animal’s state of health is less favorable than it would be in the absence of the disorder. Left untreated, a disorder does not necessarily cause a further decrease in the animal’s state of health.

As used herein, an “effective amount” means an amount sufficient to produce a selected effect. A “therapeutically effective amount” means an effective amount of an agent being used in treating or preventing a disease or disorder.

As used herein, a “functional” molecule is a molecule in a form in which it exhibits a property or activity by which it is characterized.

As used herein, a “functional biological molecule” is a biological molecule in a form in which it exhibits a property by which it is characterized. A functional enzyme, for example, is one which exhibits the characteristic catalytic activity by which the enzyme is characterized.

“Homologous” as used herein, refers to the subunit sequence similarity between two polymeric molecules, e.g., between two nucleic acid molecules, e.g., two DNA molecules or two RNA molecules, or between two polypeptide molecules. When a subunit position in both of the two molecules is occupied by the same monomeric subunit, e.g., if a position in each of two DNA molecules is occupied by adenine, then they are homologous at that position. The homology between two sequences is a direct function of the number of matching or homologous positions, e.g., if half (e.g., five positions in a polymer ten subunits in length) of the positions in two compound sequences are homologous then the two sequences are 50% homologous, if 90% of the positions, e.g., 9 of 10, are matched or homologous, the two sequences share 90% homology. By way of example, the DNA sequences 5’-ATTGCC-3’ and 5’-TATGGC-3’ share 50% homology.

As used herein, “homology” is used synonymously with “identity”.

The determination of percent identity between two nucleotide or amino acid sequences can be accomplished using a mathematical algorithm. For example, a mathematical algorithm useful for comparing two sequences is the algorithm of Karlin & Altschul, 1990, modified as in Karlin & Altschul, 1993). This algorithm is incorporated into the NBLAST and XBLAST programs (see Altschul et al., 1990a; Altschul et al., 1990b),

and can be accessed, for example at the National Center for Biotechnology Information (NCBI) world wide web site. BLAST nucleotide searches can be performed with the NBLAST program (designated “blastn” at the NCBI web site), using the following parameters: gap penalty = 5; gap extension penalty = 2; mismatch penalty = 3; match reward = 1; expectation value 10.0; and word size = 11 to obtain nucleotide sequences homologous to a nucleic acid described herein. BLAST protein searches can be performed with the XBLAST program (designated “blastn” at the NCBI web site) or the NCBI “blastp” program, using the following parameters: expectation value 10.0, BLOSUM62 scoring matrix to obtain amino acid sequences homologous to a protein molecule described herein. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al., 1997. Alternatively, PSI-Blast or PHI-Blast can be used to perform an iterated search which detects distant relationships between molecules (Id.) and relationships between molecules which share a common pattern. When utilizing BLAST, Gapped BLAST, PSI-Blast, and PHI-Blast programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used.

The percent identity between two sequences can be determined using techniques similar to those described above, with or without allowing gaps. In calculating percent identity, typically exact matches are counted.

As used herein, the term “hybridization” is used in reference to the pairing of complementary nucleic acids. Hybridization and the strength of hybridization (i.e., the strength of the association between the nucleic acids) is impacted by such factors as the degree of complementarity between the nucleic acids, stringency of the conditions involved, the length of the formed hybrid, and the G:C ratio within the nucleic acids.

The term “ingredient” refers to any compound, whether of chemical or biological origin, that can be used in cell culture media to maintain or promote the proliferation, survival, or differentiation of cells. The terms “component”, “nutrient”, “supplement”, and “ingredient” can be used interchangeably and are all meant to refer to such compounds. Typical non-limiting ingredients that are used in cell culture media include amino acids, salts, metals, sugars, lipids, nucleic acids, hormones, vitamins, fatty acids, proteins, and the like. Other ingredients that promote or maintain cultivation of cells *ex vivo* can be selected by those of skill in the art, in accordance with the particular need.

The term “inhibit”, as used herein, refers to the ability of a compound, agent, or method to reduce or impede a described function, level, activity, rate, etc., based on the

context in which the term “inhibit” is used. In some embodiments, inhibition is by at least 10%, in some embodiments by at least 25%, in some embodiments by at least 50%, and in some embodiments, the function is inhibited by at least 75%. The term “inhibit” is used interchangeably with “reduce” and “block”.

The term “inhibitor” as used herein, refers to any compound or agent, the application of which results in the inhibition of a process or function of interest, including, but not limited to, differentiation and activity. Inhibition can be inferred if there is a reduction in the activity or function of interest.

As used herein “injecting or applying” includes administration of a compound of the presently disclosed subject matter by any number of routes and means including, but not limited to, topical, oral, buccal, intravenous, intramuscular, intra arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, vaginal, ophthalmic, pulmonary, or rectal means.

As used herein, “injury” generally refers to damage, harm, or hurt; usually applied to damage inflicted on the body by an external force.

Used interchangeably herein are the terms “isolate” and “select”.

The term “isolated”, when used in reference to cells, refers to a single cell of interest, or population of cells of interest, at least partially isolated from other cell types or other cellular material with which it naturally occurs in the tissue of origin (e.g., adipose tissue). A sample of stem cells is “substantially pure” when it is in some embodiments at least 60%, in some embodiments at least 75%, in some embodiments at least 90%, and, in certain cases, in some embodiments at least 99% free of cells other than cells of interest. Purity can be measured by any appropriate method, for example, by fluorescence-activated cell sorting (FACS), or other assays, which distinguish cell types.

An “isolated nucleic acid” refers to a nucleic acid segment or fragment, which has been separated from sequences, which flank it in a naturally occurring state, e.g., a DNA fragment that has been removed from the sequences, which are normally adjacent to the fragment, e.g., the sequences adjacent to the fragment in a genome in which it naturally occurs. The term also applies to nucleic acids, which have been substantially purified, from other components, which naturally accompany the nucleic acid, e.g., RNA or DNA, or proteins, which naturally accompany it in the cell. The term therefore includes, for example, a recombinant DNA which is incorporated into a vector, into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote, or which exists as

a separate molecule (e.g., as a cDNA or a genomic or cDNA fragment produced by PCR or restriction enzyme digestion) independent of other sequences. It also includes a recombinant DNA, which is part of a hybrid gene encoding additional polypeptide sequence.

Unless otherwise specified, a “nucleotide sequence encoding an amino acid sequence” includes all nucleotide sequences that are degenerate versions of each other and that encode the same amino acid sequence. Nucleotide sequences that encode proteins and RNA may include introns.

As used herein, a “ligand” is a compound that specifically binds to a target compound. A ligand (e.g., an antibody) “specifically binds to” or “is specifically immunoreactive with” a compound when the ligand functions in a binding reaction which is determinative of the presence of the compound in a sample of heterogeneous compounds. Thus, under designated assay (e.g., immunoassay) conditions, the ligand binds preferentially to a particular compound and does not bind to a significant extent to other compounds present in the sample. For example, an antibody specifically binds under immunoassay conditions to an antigen bearing an epitope against which the antibody was raised. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular antigen. For example, solid-phase ELISA immunoassays are routinely used to select monoclonal antibodies specifically immunoreactive with an antigen. See Harlow & Lane, 1988 for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity.

A “receptor” is a compound that specifically or selectively binds to a ligand.

As used herein, the term “linkage” refers to a connection between two groups. The connection can be either covalent or non-covalent, including but not limited to ionic bonds, hydrogen bonding, and hydrophobic/hydrophilic interactions.

As used herein, the term “linker” refers to either a molecule that joins two other molecules covalently or noncovalently, e.g., through ionic or hydrogen bonds or van der Waals interactions.

The terms “gene product” or “expression product” are used herein interchangeably to refer to the RNA transcription products (RNA transcript) of a gene, including mRNA, and the polypeptide translation product of such RNA transcripts. A gene product may be, for example, a polynucleotide gene expression product (e.g., an unspliced RNA, an mRNA, a splice variant mRNA, a microRNA, a fragmented RNA, and the like) or a protein expression product (e.g., a mature polypeptide, a post-translationally modified polypeptide,

a splice variant polypeptide, and the like). In some embodiments the gene expression product may be a sequence variant including mutations, fusions, loss of heterozygosity (LOH), and/or biological pathway effects.

The term “measuring the level of expression” or “determining the level of expression” as used herein refers to any measure or assay which can be used to correlate the results of the assay with the level of expression of a gene or protein of interest. Such assays include measuring the level of mRNA, protein levels, etc. and can be performed by assays such as northern and western blot analyses, binding assays, immunoblots, etc. The level of expression can include rates of expression and can be measured in terms of the actual amount of an mRNA or protein present. Such assays are coupled with processes or systems to store and process information and to help quantify levels, signals, etc. and to digitize the information for use in comparing levels.

A “reference expression level” as applied to a gene expression product refers to an expression level for one or more reference (or “control”) gene expression products. A “reference normalized expression level” as applied to a gene expression product refers to a normalized expression level value for one or more reference (or control) gene expression products (i.e., a normalized reference expression level). In some embodiments, a reference expression level is an expression level for one or more gene product in normal sample, as described herein. In some embodiments, a reference expression level is determined experimentally. In some embodiments, a reference expression level is a historical expression level, e.g., a database value of a reference expression level in a normal sample, which sample indicates a single reference expression level, or a summary of a plurality of reference expression levels (such as, e.g., (i) an average of two or more, in some embodiments three or more reference expression levels from replicate analysis of the reference expression level from a single sample; (ii) an average of two or more, in some embodiments three or more reference expression levels from analysis of the reference expression level from a plurality of different samples (e.g., normal samples); (iii) and a combination of the above mentioned steps (i) and (ii) (i.e., average of reference expression levels analyzed from a plurality of samples, wherein at least one of the reference expression levels are analyzed in replicate). In some embodiments, the “reference expression level” is an expression level of sequence variants, for example, in a sample that has been definitively determined to be UIP or non-UIP by other approaches (i.e. confirmed pathological diagnosis).

A “reference expression level value” as applied to a gene expression product refers to an expression level value for one or more reference (or control) gene expression products. A “reference normalized expression level value” as applied to a gene expression product refers to a normalized expression level value for one or more reference (or control) gene expression products.

“Stringency” of hybridization reactions is readily determinable by one of ordinary skill in the art, and generally is an empirical calculation dependent upon probe length, washing temperature, and salt concentration. In general, longer probes require higher temperatures for proper annealing, while shorter probes need lower temperatures. Hybridization generally depends on the ability of denatured DNA to re-anneal when complementary strands are present in an environment below their melting temperature. The higher the degree of desired homology between the probe and hybridizable sequence, the higher the relative temperature that may be used. As a result, it follows that higher relative temperatures may tend to make the reaction conditions more stringent, while lower temperatures less so. For additional details and explanation of stringency of hybridization reactions, see Ausubel et al., 1995.

“Stringent conditions” or “high stringency conditions”, as defined herein, typically: (1) employ low ionic strength solutions and high temperature for washing, for example 0.015 M sodium chloride/0.0015 M sodium citrate/0.1% sodium dodecyl sulfate at 50° C.; (2) employ during hybridization a denaturing agent, such as formamide, for example, 50% (v/v) formamide with 0.1% bovine serum albumin/0.1% Ficoll/0.1% polyvinylpyrrolidone/50 mM sodium phosphate buffer at pH 6.5 with 750 mM sodium chloride, 75 mM sodium citrate at 42° C.; or (3) employ 50% formamide, 5× SSC (0.75 M NaCl, 0.075 M sodium citrate), 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, 5× Denhardt's solution, sonicated salmon sperm DNA (50 µg/ml), 0.1% SDS, and 10% dextran sulfate at 42° C., with washes at 42° C. in 0.2× SSC (sodium chloride/sodium citrate) and 50% formamide at 55° C., followed by a high-stringency wash consisting of 0.1× SSC containing EDTA at 55° C.

“Moderately stringent conditions” may be identified as described by Sambrook et al., 1989, and include the use of washing solution and hybridization conditions (e.g., temperature, ionic strength and % SDS) less stringent than those described above. An example of moderately stringent condition is overnight incubation at 37° C. in a solution comprising: 20% formamide, 5×SSC (150 mM NaCl, 15 mM trisodium citrate), 50 mM

sodium phosphate (pH 7.6), 5×Denhardt's solution, 10% dextran sulfate, and 20 mg/ml denatured sheared salmon sperm DNA, followed by washing the filters in 1×SSC at about 37-50° C. The skilled artisan will recognize how to adjust the temperature, ionic strength, etc. as necessary to accommodate factors such as probe length and the like.

“Sensitivity” as used herein refers to the proportion of true positives of the total number tested that actually have the target disorder (i.e., the proportion of patients with the target disorder who have a positive test result). “Specificity” as used herein refers to the proportion of true negatives of all the patients tested who actually do not have the target disorder (i.e., the proportion of patients without the target disorder who have a negative test result).

In the context of the present disclosure, reference to “at least one,” “at least two,” “at least five,” etc. of the genes listed in any particular gene set means any one or any and all combinations of the genes listed.

The term “modulate”, as used herein, refers to changing the level of an activity, function, or process. The term “modulate” encompasses both inhibiting and stimulating an activity, function, or process. The term “modulate” is used interchangeably with the term “regulate” herein.

The term “nucleic acid” typically refers to large polynucleotides. By “nucleic acid” is meant any nucleic acid, whether composed of deoxyribonucleosides or ribonucleosides, and whether composed of phosphodiester linkages or modified linkages such as phosphotriester, phosphoramidate, siloxane, carbonate, carboxymethylester, acetamidate, carbamate, thioether, bridged phosphoramidate, bridged methylene phosphonate, bridged phosphoramidate, bridged phosphoramidate, bridged methylene phosphonate, phosphorothioate, methylphosphonate, phosphorodithioate, bridged phosphorothioate or sulfone linkages, and combinations of such linkages. The term nucleic acid also specifically includes nucleic acids composed of bases other than the five biologically occurring bases (adenine, guanine, thymine, cytosine, and uracil).

As used herein, the term “nucleic acid” encompasses RNA as well as single and double stranded DNA and cDNA. Furthermore, the terms, “nucleic acid”, “DNA”, “RNA” and similar terms also include nucleic acid analogs, i.e. analogs having other than a phosphodiester backbone. For example, the so called “peptide nucleic acids”, which are known in the art and have peptide bonds instead of phosphodiester bonds in the backbone, are considered within the scope of the presently disclosed subject matter. By “nucleic acid”

is meant any nucleic acid, whether composed of deoxyribonucleosides or ribonucleosides, and whether composed of phosphodiester linkages or modified linkages such as phosphotriester, phosphoramidate, siloxane, carbonate, carboxymethylester, acetamidate, carbamate, thioether, bridged phosphoramidate, bridged methylene phosphonate, bridged phosphoramidate, bridged phosphoramidate, bridged methylene phosphonate, phosphorothioate, methylphosphonate, phosphorodithioate, bridged phosphorothioate or sulfone linkages, and combinations of such linkages. The term nucleic acid also specifically includes nucleic acids composed of bases other than the five biologically occurring bases (adenine, guanine, thymine, cytosine, and uracil). Conventional notation is used herein to describe polynucleotide sequences: the left-hand end of a single-stranded polynucleotide sequence is the 5'-end; the left-hand direction of a double-stranded polynucleotide sequence is referred to as the 5'-direction. The direction of 5' to 3' addition of nucleotides to nascent RNA transcripts is referred to as the transcription direction. The DNA strand having the same sequence as an mRNA is referred to as the "coding strand"; sequences on the DNA strand which are located 5' to a reference point on the DNA are referred to as "upstream sequences"; sequences on the DNA strand which are 3' to a reference point on the DNA are referred to as "downstream sequences".

The term "nucleic acid construct", as used herein, encompasses DNA and RNA sequences encoding the particular gene or gene fragment desired, whether obtained by genomic or synthetic methods.

Unless otherwise specified, a "nucleotide sequence encoding an amino acid sequence" includes all nucleotide sequences that are degenerate versions of each other and that encode the same amino acid sequence. Nucleotide sequences that encode proteins and RNA may include introns.

The term "oligonucleotide" typically refers to short polynucleotides, generally, no greater than about 50 nucleotides. It will be understood that when a nucleotide sequence is represented by a DNA sequence (i.e., A, T, G, C), this also includes an RNA sequence (i.e., A, U, G, C) in which "U" replaces "T".

By describing two polynucleotides as "operably linked" is meant that a single-stranded or double-stranded nucleic acid moiety comprises the two polynucleotides arranged within the nucleic acid moiety in such a manner that at least one of the two polynucleotides is able to exert a physiological effect by which it is characterized upon the

other. By way of example, a promoter operably linked to the coding region of a gene is able to promote transcription of the coding region.

As used herein, “parenteral administration” of a pharmaceutical composition includes any route of administration characterized by physical breaching of a tissue of a subject and administration of the pharmaceutical composition through the breach in the tissue. Parenteral administration thus includes, but is not limited to, administration of a pharmaceutical composition by injection of the composition, by application of the composition through a surgical incision, by application of the composition through a tissue-penetrating non-surgical wound, and the like. In particular, parenteral administration is contemplated to include, but is not limited to, subcutaneous, intraperitoneal, intramuscular, intrasternal injection, and kidney dialytic infusion techniques.

The term “pharmaceutical composition” shall mean a composition comprising at least one active ingredient, whereby the composition is amenable to investigation for a specified, efficacious outcome in a mammal (for example, without limitation, a human). Those of ordinary skill in the art will understand and appreciate the techniques appropriate for determining whether an active ingredient has a desired efficacious outcome based upon the needs of the artisan.

As used herein, the term “pharmaceutically-acceptable carrier” means a chemical composition with which an appropriate compound or derivative can be combined and which, following the combination, can be used to administer the appropriate compound to a subject.

As used herein, the term “physiologically acceptable” ester or salt means an ester or salt form of the active ingredient which is compatible with any other ingredients of the pharmaceutical composition, which is not deleterious to the subject to which the composition is to be administered.

“Plurality” means at least two.

A “polynucleotide” means a single strand or parallel and anti-parallel strands of a nucleic acid. Thus, a polynucleotide may be either a single-stranded or a double-stranded nucleic acid.

“Polypeptide” refers to a polymer composed of amino acid residues, related naturally occurring structural variants, and synthetic non-naturally occurring analogs thereof linked via peptide bonds, related naturally occurring structural variants, and synthetic non-naturally occurring analogs thereof.

“Synthetic peptides or polypeptides” means a non-naturally occurring peptide or polypeptide. Synthetic peptides or polypeptides can be synthesized, for example, using an automated polypeptide synthesizer. Various solid phase peptide synthesis methods are known to those of skill in the art.

The term “prevent”, as used herein, means to stop something from happening, or taking advance measures against something possible or probable from happening. In the context of medicine, “prevention” generally refers to action taken to decrease the chance of getting a disease or condition.

“Primer” refers to a polynucleotide that is capable of specifically hybridizing to a designated polynucleotide template and providing a point of initiation for synthesis of a complementary polynucleotide. Such synthesis occurs when the polynucleotide primer is placed under conditions in which synthesis is induced, i.e., in the presence of nucleotides, a complementary polynucleotide template, and an agent for polymerization such as DNA polymerase. A primer is typically single-stranded, but may be double-stranded. Primers are typically deoxyribonucleic acids, but a wide variety of synthetic and naturally occurring primers are useful for many applications. A primer is complementary to the template to which it is designed to hybridize to serve as a site for the initiation of synthesis, but need not reflect the exact sequence of the template. In such a case, specific hybridization of the primer to the template depends on the stringency of the hybridization conditions. Primers can be labeled with, e.g., chromogenic, radioactive, or fluorescent moieties and used as detectable moieties.

A “prophylactic” treatment is a treatment administered to a subject who does not exhibit signs of a disease or injury or exhibits only early signs of the disease or injury for the purpose of decreasing the risk of developing pathology associated with the disease or injury.

As used herein, “protecting group” with respect to a terminal amino group refers to a terminal amino group of a peptide, which terminal amino group is coupled with any of various amino-terminal protecting groups traditionally employed in peptide synthesis. Such protecting groups include, for example, acyl protecting groups such as formyl, acetyl, benzoyl, trifluoroacetyl, succinyl, and methoxysuccinyl; aromatic urethane protecting groups such as benzyloxycarbonyl; and aliphatic urethane protecting groups, for example, tert-butoxycarbonyl or adamantyloxycarbonyl. See Gross & Mienhofer, 1981 for suitable protecting groups.

As used herein, “protecting group” with respect to a terminal carboxy group refers to a terminal carboxyl group of a peptide, which terminal carboxyl group is coupled with any of various carboxyl-terminal protecting groups. Such protecting groups include, for example, tert-butyl, benzyl, or other acceptable groups linked to the terminal carboxyl group through an ester or ether bond.

The term “protein” typically refers to large polypeptides. Conventional notation is used herein to portray polypeptide sequences: the left-hand end of a polypeptide sequence is the amino-terminus; the right-hand end of a polypeptide sequence is the carboxyl-terminus.

The term “protein regulatory pathway”, as used herein, refers to both the upstream regulatory pathway which regulates a protein, as well as the downstream events which that protein regulates. Such regulation includes, but is not limited to, transcription, translation, levels, activity, posttranslational modification, and function of the protein of interest, as well as the downstream events which the protein regulates.

The terms “protein pathway” and “protein regulatory pathway” are used interchangeably herein.

As used herein, the term “purified” and like terms relate to an enrichment of a molecule or compound relative to other components normally associated with the molecule or compound in a native environment. The term “purified” does not necessarily indicate that complete purity of the particular molecule has been achieved during the process. A “highly purified” compound as used herein refers to a compound that is greater than 90% pure.

“Recombinant polynucleotide” refers to a polynucleotide having sequences that are not naturally joined together. An amplified or assembled recombinant polynucleotide may be included in a suitable vector, and the vector can be used to transform a suitable host cell.

A recombinant polynucleotide may serve a non-coding function (e.g., promoter, origin of replication, ribosome-binding site, etc.) as well.

A host cell that comprises a recombinant polynucleotide is referred to as a “recombinant host cell”. A gene which is expressed in a recombinant host cell wherein the gene comprises a recombinant polynucleotide, produces a “recombinant polypeptide”.

A “recombinant polypeptide” is one which is produced upon expression of a recombinant polynucleotide.

The term “regulate” refers to either stimulating or inhibiting a function or activity of interest.

As used herein, term “regulatory elements” is used interchangeably with “regulatory sequences” and refers to promoters, enhancers, and other expression control elements, or any combination of such elements.

A “sample”, as used herein, refers in some embodiments to a biological sample from a subject, including, but not limited to, normal tissue samples, diseased tissue samples, biopsies, blood, saliva, feces, semen, tears, and urine. A sample can also be any other source of material obtained from a subject which contains cells, tissues, or fluid of interest. A sample can also be obtained from cell or tissue culture.

A “significant detectable level” is an amount of contaminate that would be visible in the presented data and would need to be addressed/explained during analysis of the forensic evidence.

By the term “signal sequence” is meant a polynucleotide sequence which encodes a peptide that directs the path a polypeptide takes within a cell, i.e., it directs the cellular processing of a polypeptide in a cell, including, but not limited to, eventual secretion of a polypeptide from a cell. A signal sequence is a sequence of amino acids which are typically, but not exclusively, found at the amino terminus of a polypeptide which targets the synthesis of the polypeptide to the endoplasmic reticulum. In some instances, the signal peptide is proteolytically removed from the polypeptide and is thus absent from the mature protein.

By “small interfering RNAs (siRNAs)” is meant, inter alia, an isolated dsRNA molecule comprised of both a sense and an anti-sense strand. In some embodiments, it is greater than 10 nucleotides in length. siRNA also refers to a single transcript which has both the sense and complementary antisense sequences from the target gene, e.g., a hairpin. siRNA further includes any form of dsRNA (proteolytically cleaved products of larger dsRNA, partially purified RNA, essentially pure RNA, synthetic RNA, recombinantly produced RNA) as well as altered RNA that differs from naturally occurring RNA by the addition, deletion, substitution, and/or alteration of one or more nucleotides.

The terms “solid support”, “surface” and “substrate” are used interchangeably and refer to a structural unit of any size, where said structural unit or substrate has a surface suitable for immobilization of molecular structure or modification of said structure and said substrate is made of a material such as, but not limited to, metal, metal films, glass, fused silica, synthetic polymers, and membranes.

By the term “specifically binds”, as used herein, is meant a molecule which recognizes and binds a specific molecule, but does not substantially recognize or bind other

molecules in a sample, or it means binding between two or more molecules as in part of a cellular regulatory process, where said molecules do not substantially recognize or bind other molecules in a sample.

The term “standard”, as used herein, refers to something used for comparison. For example, it can be a known standard agent or compound which is administered and used for comparing results when administering a test compound, or it can be a standard parameter or function which is measured to obtain a control value when measuring an effect of an agent or compound on a parameter or function. “Standard” can also refer to an “internal standard”, such as an agent or compound which is added at known amounts to a sample and which is useful in determining such things as purification or recovery rates when a sample is processed or subjected to purification or extraction procedures before a marker of interest is measured. Internal standards are often but are not limited to, a purified marker of interest which has been labeled, such as with a radioactive isotope, allowing it to be distinguished from an endogenous substance in a sample.

The term “stimulate” as used herein, means to induce or increase an activity or function level such that it is higher relative to a control value. The stimulation can be via direct or indirect mechanisms. In some embodiments, the activity or function is stimulated by at least 10% compared to a control value, in some embodiments by at least 25%, and in some embodiments by at least 50%. The term “stimulator” as used herein, refers to any composition, compound or agent, the application of which results in the stimulation of a process or function of interest, including, but not limited to, wound healing, angiogenesis, bone healing, osteoblast production and function, and osteoclast production, differentiation, and activity.

The term “subject,” as used herein, generally refers to a mammal. Typically, the subject is a human. However, the term embraces other species, e.g., pigs, mice, rats, dogs, cats, or other primates. In certain embodiments, the subject is an experimental subject such as a mouse or rat. The subject may be a male or female. The subject may be an infant, a toddler, a child, a young adult, an adult or a geriatric. The subject may exhibit one or more symptoms of IPF. For example, the subject may exhibit shortness of breath (generally aggravated by exertion) and/or dry cough), and, in some cases may have obtained results of one or more of an imaging test (e.g., chest X-ray, computerized tomography (CT)), a pulmonary function test (e.g., spirometry, oximetry, exercise stress test), lung tissue analysis (e.g., histological and/or cytological analysis of samples obtained by bronchoscopy,

bronchoalveolar lavage, surgical biopsy) that is indicative of the potential presence of IPF. A subject under the care of a physician or other health care provider may be referred to as a “patient”.

A “subject” of diagnosis or treatment is an animal, including a human. It also includes pets and livestock.

As used herein, a “subject in need thereof” is a patient, animal, mammal, or human, who will benefit from the method of the presently disclosed subject matter.

As used herein, “substantially homologous amino acid sequences” includes those amino acid sequences which have at least about 95% homology, in some embodiments at least about 96% homology, more in some embodiments at least about 97% homology, in some embodiments at least about 98% homology, and most in some embodiments at least about 99% or more homology to an amino acid sequence of a reference sequence. Amino acid sequence similarity or identity can be computed by using the BLASTP and TBLASTN programs which employ the BLAST (basic local alignment search tool) 2.0.14 algorithm. The default settings used for these programs are suitable for identifying substantially similar amino acid sequences for purposes of the presently disclosed subject matter.

“Substantially homologous nucleic acid sequence” means a nucleic acid sequence corresponding to a reference nucleic acid sequence wherein the corresponding sequence encodes a peptide having substantially the same structure and function as the peptide encoded by the reference nucleic acid sequence; e.g., where only changes in amino acids not significantly affecting the peptide function occur. In some embodiments, the substantially identical nucleic acid sequence encodes the peptide encoded by the reference nucleic acid sequence. The percentage of identity between the substantially similar nucleic acid sequence and the reference nucleic acid sequence is at least about 50%, 65%, 75%, 85%, 95%, 99% or more. Substantial identity of nucleic acid sequences can be determined by comparing the sequence identity of two sequences, for example by physical/chemical methods (i.e., hybridization) or by sequence alignment via computer algorithm. Suitable nucleic acid hybridization conditions to determine if a nucleotide sequence is substantially similar to a reference nucleotide sequence are: 7% sodium dodecyl sulfate SDS, 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 2X standard saline citrate (SSC), 0.1% SDS at 50°C; in some embodiments in 7% (SDS), 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 1X SSC, 0.1% SDS at 50°C; in some embodiments 7% SDS, 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 0.5X SSC, 0.1% SDS at 50°C; and more in some

embodiments in 7% SDS, 0.5 M NaPO₄, 1 mM EDTA at 50°C with washing in 0.1X SSC, 0.1% SDS at 65°C. Suitable computer algorithms to determine substantial similarity between two nucleic acid sequences include, GCS program package (Devereux et al., 1984), and the BLASTN or FASTA programs (Altschul et al., 1990a; Altschul et al., 1990b; Altschul et al., 1997). The default settings provided with these programs are suitable for determining substantial similarity of nucleic acid sequences for purposes of the presently disclosed subject matter.

The term “substantially pure” describes a compound, e.g., a protein or polypeptide which has been separated from components which naturally accompany it. Typically, a compound is substantially pure when at least 10%, more in some embodiments at least 20%, more in some embodiments at least 50%, more in some embodiments at least 60%, more in some embodiments at least 75%, more in some embodiments at least 90%, and most in some embodiments at least 99% of the total material (by volume, by wet or dry weight, or by mole percent or mole fraction) in a sample is the compound of interest. Purity can be measured by any appropriate method, e.g., in the case of polypeptides by column chromatography, gel electrophoresis, or HPLC analysis. A compound, e.g., a protein, is also substantially purified when it is essentially free of naturally associated components or when it is separated from the native contaminants which accompany it in its natural state.

A “surface active agent” or “surfactant” is a substance that has the ability to reduce the surface tension of materials and enable penetration into and through materials.

The term “symptom”, as used herein, refers to any morbid phenomenon or departure from the normal in structure, function, or sensation, experienced by the patient and indicative of disease. In contrast, a “sign” is objective evidence of disease. For example, a bloody nose is a sign. It is evident to the patient, doctor, nurse, and other observers.

A “therapeutic” treatment is a treatment administered to a subject who exhibits signs of pathology for the purpose of diminishing or eliminating those signs.

A “therapeutically effective amount” of a compound is that amount of compound which is sufficient to provide a beneficial effect to the subject to which the compound is administered.

“Tissue” means (1) a group of similar cell united perform a specific function; (2) a part of an organism consisting of an aggregate of cells having a similar structure and function; or (3) a grouping of cells that are similarly characterized by their structure and function, such as muscle or nerve tissue.

The term “topical application”, as used herein, refers to administration to a surface, such as the skin. This term is used interchangeably with “cutaneous application” in the case of skin. A “topical application” is a “direct application”.

By “transdermal” delivery is meant delivery by passage of a drug through the skin or mucosal tissue and into the bloodstream. Transdermal also refers to the skin as a portal for the administration of drugs or compounds by topical application of the drug or compound thereto. “Transdermal” is used interchangeably with “percutaneous”.

The term “transfection” is used interchangeably with the terms “gene transfer”, “transformation”, and “transduction”, and means the intracellular introduction of a polynucleotide. “Transfection efficiency” refers to the relative amount of the transgene taken up by the cells subjected to transfection. In practice, transfection efficiency is estimated by the amount of the reporter gene product expressed following the transfection procedure.

As used herein, the term “transgene” means an exogenous nucleic acid sequence comprising a nucleic acid which encodes a promoter/regulatory sequence operably linked to nucleic acid which encodes an amino acid sequence, which exogenous nucleic acid is encoded by a transgenic mammal.

As used herein, the term “treating” may include prophylaxis of the specific injury, disease, disorder, or condition, or alleviation of the symptoms associated with a specific injury, disease, disorder, or condition and/or preventing or eliminating said symptoms. A “prophylactic” treatment is a treatment administered to a subject who does not exhibit signs of a disease or exhibits only early signs of the disease for the purpose of decreasing the risk of developing pathology associated with the disease. “Treating” is used interchangeably with “treatment” herein.

A “vector” is a composition of matter which comprises an isolated nucleic acid and which can be used to deliver the isolated nucleic acid to the interior of a cell. Numerous vectors are known in the art including, but not limited to, linear polynucleotides, polynucleotides associated with ionic or amphiphilic compounds, plasmids, and viruses. Thus, the term “vector” includes an autonomously replicating plasmid or a virus. The term should also be construed to include non-plasmid and non-viral compounds which facilitate transfer or delivery of nucleic acid to cells, such as, for example, polylysine compounds, liposomes, and the like. Examples of viral vectors include, but are not limited to, adenoviral vectors, adeno-associated virus vectors, retroviral vectors, recombinant viral vectors, and

the like. Examples of non-viral vectors include, but are not limited to, liposomes, polyamine derivatives of DNA and the like.

“Expression vector” refers to a vector comprising a recombinant polynucleotide comprising expression control sequences operatively linked to a nucleotide sequence to be expressed. An expression vector comprises sufficient cis-acting elements for expression; other elements for expression can be supplied by the host cell or in an in vitro expression system. Expression vectors include all those known in the art, such as cosmids, plasmids (e.g., naked or contained in liposomes) and viruses that incorporate the recombinant polynucleotide.

As used herein “wound” or “wounds” may refer to any detectable break in the tissues of the body, such as injury to skin or to an injury or damage, or to a damaged site associated with a disease or disorder. As used herein, the term “wound” relates to a physical tear, break, or rupture to a tissue or cell layer. A wound may occur by any physical insult, including a surgical procedure or as a result of a disease, disorder condition. Although the terms “wound” and “injury” are not always defined exactly the same way, the use of one term herein, such as “injury”, is not meant to exclude the meaning of the other term.

III. Methods and Uses of the Presently Disclosed Subject Matter

The presently disclosed subject matter relates in some embodiments to methods for identifying, classifying, and treating patients with Idiopathic Pulmonary Fibrosis (IPF) as suffering from or being at risk for developing a longitudinal decline in forced vital capacity (FVC).

As used herein, the phrase “forced vital capacity” (FVC) refers to that amount of air that can be forcibly exhaled from the lungs after taking one’s deepest breath. FVC is typically measured by spirometry. It can be employed to distinguish between obstructive versus restrictive lung diseases. In Idiopathic Pulmonary Fibrosis (IPF), a longitudinal decline in FVC is a well-validated predictor of mortality and is often used as the primary efficacy endpoint in IPF clinical trials. With respect to FVC decline, IPF patients can be categorized as being stable (referred to herein as FVC-stable or FVC-S) or can have and/or be at risk for progressive disease. In some embodiments, an FVC-S patient is a patient who would not be predicted to suffer a $\geq 10\%$ relative decline in FVC over the next 12 months. In some embodiments, an FVC-S patient is a patient who would not be predicted to suffer a $\geq 5\%$ relative decline in FVC over the next 12 months. Thus, in some embodiments, as used herein, progressive disease is defined as a $\geq 10\%$ relative decline in FVC over the next 12

months, and in some embodiments progressive disease is defined as a $\geq 5\%$ relative decline in FVC over the next 12 months. Accordingly, an FVC-D patient is a patient who the presently disclosed methods would be predicted to suffer a $\geq 10\%$ relative decline in FVC over the next 12 months, and in some embodiments an FVC-D patient is a patient who would be predicted to suffer a $\geq 5\%$ relative decline in FVC over the next 12 months. As used herein, the term “decline” as employed in the context of FVC is synonymous with the term “progressor”.

As disclosed in more detail herein below, the methods of the methods of the presently disclosed subject matter can be employed to identify, classify, and treat IPF patients suffering from and/or being at risk for developing progressive disease as defined herein as a longitudinal decline in FVC if in some embodiments $\geq 5\%$ and in some embodiments $\geq 10\%$ in the 12 months subsequent to testing.

III.A. Methods for Generating Prognostic Signatures

In order to identify patients (in some embodiments, human patients) suffering from or at risk for suffering progressive disease, in some embodiments the presently disclosed subject matter provides methods for generating prognostic signatures for IPF subjects with respect to a decline in FVC. In some embodiments, the methods comprise performing gene expression analysis with respect to one or more of the gene products disclosed herein at an initial and at a subsequent timepoint and comparing the first and second expression levels for the one or more genes, wherein the comparing provides a prognostic signature for the subject with respect to decline in lung FVC within a pre-determined time period subsequent to the later timepoint.

In some embodiments, the initial timepoint serves as to provide baseline values for the expression levels of the genes for the patient. In some embodiments, the subsequent timepoint provides later gene expression values for the patient, and when compared to the initial baseline values, can provide a prognostic signature that predicts whether or not the patient is likely to suffer from progressive disease within a pre-determined time period subsequent to the subsequent timepoint. In some embodiments, the initial (e.g., first) and subsequent (e.g., second) timepoints are separated by one or several months, which in some embodiments can be from about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or about 12 months.

As disclosed herein, various genes have been identified as being relevant to the production of a prognostic signature with respect to patients suffering from or at risk for suffering progressive disease. These genes are listed below in Table 1.

Table 1

Exemplary Human Genes Employed in Generating Prognostic Signatures

GENE	Description	Nucleic Acid*	Amino Acid**
<i>ALDH4A1</i>	aldehyde dehydrogenase 4 family member A1	NM_003748.4 (SEQ ID NO: 1)	NP_003739.2 (SEQ ID NO: 2)
<i>APTX</i>	Aprataxin	NM_175073.2 (SEQ ID NO: 3)	NP_778243.1 (SEQ ID NO: 4)
<i>ATP6AP1L</i>	ATPase H ⁺ transporting accessory protein 1 like	NM_001349372.2 (SEQ ID NO: 5)	NP_001336301.1 (SEQ ID NO: 6)
<i>CCNB1</i>	cyclin B1	NM_031966.4 (SEQ ID NO: 7)	NP_114172.1 (SEQ ID NO: 8)
<i>CNR2</i>	Cannabinoid receptor 2	NM_001841.3 (SEQ ID NO: 9)	NP_001832.1 (SEQ ID NO: 10)
<i>DNAJC17</i>	DnaJ heat shock protein family (Hsp40) member C17	NM_018163.3 (SEQ ID NO: 11)	NP_060633.1 (SEQ ID NO: 12)
<i>DTWD1</i>	DTW domain containing 1	NM_020234.6 (SEQ ID NO: 13)	NP_064619.2 (SEQ ID NO: 14)
<i>FAM111B</i>	Family with sequence similarity 111 member B	NM_198947.4 (SEQ ID NO: 15)	NP_945185.1 (SEQ ID NO: 16)
<i>GABRR1</i>	γ -aminobutyric acid type A receptor rho1 subunit	NM_002042.5 (SEQ ID NO: 17)	NP_002033.2 (SEQ ID NO: 18)
<i>GPR39</i>	G protein-coupled receptor 39	NM_001508.3 (SEQ ID NO: 19)	NP_001499.1 (SEQ ID NO: 20)
<i>GYPA</i>	Glycophorin A (MNS blood group)	NM_002099.8 (SEQ ID NO: 21)	NP_002090.4 (SEQ ID NO: 22)
<i>HBB</i>	Hemoglobin subunit beta	NM_000518.5 (SEQ ID NO: 23)	NP_000509.1 (SEQ ID NO: 24)
<i>HLA-DPBI</i>	major histocompatibility complex, class II, DP beta 1	NM_002121.6 (SEQ ID NO: 25)	NP_002112.3 (SEQ ID NO: 26)
<i>IGLC1</i>	Immunoglobulin lambda constant 1	BC012159.1 (SEQ ID NO: 27)	AAH12159.1 (SEQ ID NO: 28)

<i>ITLN1</i>	Intelectin 1	NM_017625.3 (SEQ ID NO: 29)	NP_060095.2 (SEQ ID NO: 30)
<i>LINC00319</i>	Long intergenic non-protein coding RNA 319	NR_152722.1 (SEQ ID NO: 31)	N/A
<i>MAZ</i>	MYC associated zinc finger protein	NM_002383.4 (SEQ ID NO: 32)	NP_002374.2 (SEQ ID NO: 33)
<i>MRPL35</i>	mitochondrial ribosomal protein L35	NM_016622.4 (SEQ ID NO: 34)	NP_057706.2 (SEQ ID NO: 35)
<i>MSR1</i>	Macrophage scavenger receptor 1	NM_138715.3 (SEQ ID NO: 36)	NP_619729.1 (SEQ ID NO: 37)
<i>NT5E</i>	5'-nucleotidase ecto	NM_002526.4 (SEQ ID NO: 38)	NP_002517.1 (SEQ ID NO: 39)
<i>PAWR</i>	Pro-apoptotic WT1 regulator	NM_002583.4 (SEQ ID NO: 40)	NP_002574.2 (SEQ ID NO: 41)
<i>PCDHB15</i>	Protocadherin beta 15	NM_018935.4 (SEQ ID NO: 42)	NP_061758.1 (SEQ ID NO: 43)
<i>PLA2G4A</i>	Phospholipase A2 group IVA	NM_024420.2 (SEQ ID NO: 44)	NP_077734.1 (SEQ ID NO: 45)
<i>PLCL1</i>	Phospholipase C like 1 (inactive)	NM_006226.4 (SEQ ID NO: 46)	NP_006217.3 (SEQ ID NO: 47)
<i>PNMA5</i>	PNMA family member 5	NM_001103151.1 (SEQ ID NO: 48)	NP_001096621.1 (SEQ ID NO: 49)
<i>RAB3C</i>	RAB3C, RAS oncogene family	NM_138453.4 (SEQ ID NO: 50)	NP_612462.1 (SEQ ID NO: 51)
<i>RBM43</i>	RNA binding motif protein 43	NM_198557.3 (SEQ ID NO: 52)	NP_940959.1 (SEQ ID NO: 53)
<i>RLBP1</i>	Retinaldehyde binding protein 1	NM_000326.5 (SEQ ID NO: 54)	NP_000317.1 (SEQ ID NO: 55)
<i>SESN3</i>	sestrin 3	NM_144665.4 (SEQ ID NO: 56)	NP_653266.2 (SEQ ID NO: 57)
<i>SLC25A37</i>	solute carrier family 25 member 37	NM_016612.4 (SEQ ID NO: 58)	NP_057696.2 (SEQ ID NO: 59)

<i>SSU72P8</i>	SSU72 pseudogene 8	NG_012760.1	N/A
<i>TP63</i>	Tumor protein p63	NM_003722.5 (SEQ ID NO: 60)	NP_003713.3 (SEQ ID NO: 61)
<i>WDR17</i>	WD repeat domain 17	NM_170710.5 (SEQ ID NO: 62)	NP_733828.2 (SEQ ID NO: 63)
<i>ZNF252P</i>	Zinc finger protein 252, pseudogene	NR_023392.1 (SEQ ID NO: 64)	N/A
<i>ZNF582</i>	zinc finger protein 582	NM_001320371.2 (SEQ ID NO: 65)	NP_001307300.1 (SEQ ID NO: 66)

* Accession No. for a representative human nucleic acid gene product in the GENBANK® biosequence database.

** Accession No. for a representative human amino acid gene product in the GENBANK® biosequence database that is encoded by the corresponding nucleic acid Accession No. It is noted that certain nucleic acid gene products are non-coding, and they are listed identified in this column with N/A (not applicable).

In some embodiments, the presently disclosed methods comprising determining a first expression level for one or more genes selected from the group consisting of *ALDH4A1*, *APTX*, *ATP6AP1L*, *CCNB1*, *CNR2*, *DNAJC17*, *DTWD1*, *FAM111B*, *GABRR1*, *GPR39*, *GYP A*, *HBB*, *HLA-DPBI*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MRPL35*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PLCL1*, *PNMA5*, *RAB3C*, *RBM43*, *RLBP1*, *SESN3*, *SLC25A37*, *SSU72P8*, *TP63*, *WDR17*, *ZNF252P*, and *ZNF582* in a first biological sample obtained from the subject diagnosed with IPF to establish a baseline expression level for the one or more genes; determining a second expression level for the same one or more genes in a second biological sample obtained from the subject, wherein the first and second biological samples comprise peripheral blood mononuclear cells (PBMCs) and/or nucleic acids extracted from PBMCs; and comparing the first and second expression levels for the one or more genes, wherein the comparing provides a prognostic signature for the subject with respect to decline in lung FVC within a pre-determined timeframe (e.g., 12 months) from the time that the first biological sample was obtained from the subject. In some embodiments, the presently disclosed methods comprising determining first and second expression levels for the genes *APTX*, *CNR2*, *GYP A*, *ITLN1*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PLA2G4A*, and *PNMA5*. In some

embodiments, the presently disclosed methods comprising determining first and second expression levels for the genes *APTX*, *ATP6AP1L*, *ITLN1*, *LINC00319*, *MAZ*, *MSR1*, *NT5E*, *PCDHB15*, *RAB3C*, *SSU72P8*, and *TP62*. In some embodiments, the presently disclosed methods comprising determining first and second expression levels for the genes *APTX*, *CNR2*, *GABRR1*, *GPR39*, *GYPA*, *HBB*, *ITLN1*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PNMA5*, *RLBP1*, and *SSU72P8*. In some embodiments, the presently disclosed methods comprising determining first and second expression levels for the genes *APTX*, *ATP6AP1L*, *CNR2*, *FAM111B*, *GABRR1*, *GPR39*, *GYPA*, *HBB*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PLCL1*, *PNMA5*, *RAB3C*, *RBM43*, *RLBP1*, *SSU72P8*, *TP63*, and *ZNF252P*.

Any technique that permits determination of a level of expression of a gene can be employed within the methods of the presently disclosed subject matter. In some embodiments, the gene expression levels for the selected genes are determined by employing a technique selected from the group consisting of RNA-seq analysis, quantitative polymerase chain reaction (PCR) including quantitative reverse transcription PCR (qRT-PCR), the use of a nucleic acid or protein array, or any combination thereof. In some embodiments, assaying the expression level is accomplished using RT-PCR, nucleic acid microarray hybridization, RNASeq, or a combination thereof. In some embodiments, the expression level is assayed by detecting a nucleotide expressed in the test sample or synthesized from a nucleotide expressed in the test sample. In some embodiments, the method comprises synthesizing cDNA from RNA expressed in the test sample prior to assaying the expression level. In some embodiments, the method comprises synthesizing double-stranded cDNA from the cDNA prior to assaying the expression level. In some embodiments, the method comprises synthesizing non-natural RNA from the double-stranded cDNA prior to assaying the expression level. In some embodiments, the non-natural RNA is cRNA. In some embodiments, the non-natural RNA is labeled. In some embodiments, the label comprises a sequencing adaptor or a biotin molecule. In some embodiments, the method comprises amplification of the nucleotide prior to assaying the expression level. Techniques for assaying gene expression levels using RT-PCR, nucleic acid and/or protein microarray hybridization, and RNA-Seq are known in the art (see e.g., U.S. Patent Nos. 5,800,992; 6,004,755; 6,013,449; 6,020,135; 6,033,860; 6,040,138; 6,177,248; 6,251,601; 6,309,822; 7,824,856; 9,920,367; 10,227,584; each of which is incorporated by reference in its entirety. See also U.S. Patent Application Publication Nos.

2010/0120097; 2011/0189679; 2014/0113333; 2015/0307874; each of which is incorporated by reference in its entirety. See also Mortazaavi et al., 2008.

In some embodiments, the presently disclosed methods comprise comparing a normalized expression level for each gene in the first biological sample to a normalized expression level for each gene in the second biological sample to generate a fold-increase and/or a fold-decrease in the second biological sample relative to the first biological sample for each gene. As used herein, the phrase “normalized expression level” as applied to a gene expression product refers in some embodiments to a level of the gene product normalized relative to one or more reference (or control) gene expression products. Exemplary reference gene expression products include the so-called “housekeeping genes”, which are genes for which expression does not vary significantly over time, with respect to different cell types, and/or under different disease conditions. Prototypical reference genes include, but are not limited to glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and β -actin. In some embodiments, an average value within an individual cohort is employed as a normalization metric, such that fold increase and fold decrease values are expressed relative to that average.

Normalized gene expression data from two different samples can be compared to each other to determine changes in gene expression between the two different samples. In some embodiments, gene expression changes are calculated as “fold differences” between the samples. Fold differences include both fold increases (which can in some embodiments be expressed as a positive number) and fold decreases (which can in some embodiments be expressed as a negative number rather than as a fractional number between 0 and 1).

A “gene signature” of a “prognostic signature” is a gene expression pattern (i.e., expression levels of one or more genes) that is indicative of some characteristic or phenotype (such as but not limited to FVC decline within a pre-determined time period). In some embodiments, a prognostic signature refers to the expression (and/or lack of expression) of a gene, a plurality of genes, a fragment of a gene or a plurality fragments of one or more genes, which expression and/or lack of expression is indicative of status of a subject as being FVC-S or as being FVC-D.

The prognostic signature can thus be in some embodiments an overall depiction of all genes assayed or, in some embodiments, a depiction of a subset of genes (such as but not limited to informative genes).

Various other software and/or hardware modules or processes may be implemented. In certain methods, feature selection and model estimation may be performed by logistic regression with lasso penalty using glmnet (Friedman et al. 2010). Raw reads may be aligned using TopHat (Trapnell et al., 2009). Gene counts may be obtained using HTSeq (Anders et al., 2014) and normalized using DESeq (Love et al., 2014). In methods, top features (N ranging from 10 to 200) were used to train a linear support vector machine (SVM; Suykens & Vandewalle, 1999) using the e1071 library (Meyer, 2014). Confidence intervals may be computed using the pROC package (Robin et al., 2011).

In addition, data may be filtered to remove data that may be considered suspect. In some embodiments, data deriving from microarray probes that have fewer than about 4, 5, 6, 7 or 8 guanosine and cytosine nucleotides may be considered to be unreliable due to their aberrant hybridization propensity or secondary structure issues. Similarly, data deriving from microarray probes that have more than about 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, or 22 guanosine and cytosine nucleotides may be considered unreliable due to their aberrant hybridization propensity or secondary structure issues.

In some cases, unreliable probe sets may be selected for exclusion from data analysis by ranking probe-set reliability against a series of reference datasets. For example, RefSeq or Ensembl (EMBL) are considered very high quality reference datasets. Data from probe sets matching RefSeq or Ensembl sequences may in some cases be specifically included in microarray analysis experiments due to their expected high reliability. Similarly data from probe-sets matching less reliable reference datasets may be excluded from further analysis, or considered on a case by case basis for inclusion. In some cases, the Ensembl high throughput cDNA (HTC) and/or mRNA reference datasets may be used to determine the probe-set reliability separately or together. In other cases, probe-set reliability may be ranked. For example, probes and/or probe-sets that match perfectly to all reference datasets such as for example RefSeq, HTC, HTSeq, and mRNA, may be ranked as most reliable (1). Furthermore, probes and/or probe-sets that match two out of three reference datasets may be ranked as next most reliable (2), probes and/or probe-sets that match one out of three reference datasets may be ranked next (3) and probes and/or probe sets that match no reference datasets may be ranked last (4). Probes and or probe-sets may then be included or excluded from analysis based on their ranking. For example, one may choose to include data from category 1, 2, 3, and 4 probe-sets; category 1, 2, and 3 probe-sets; category 1 and 2 probe-sets; or category 1 probe-sets for further analysis. In another example, probe-sets may

be ranked by the number of base pair mismatches to reference dataset entries. It is understood that there are many methods understood in the art for assessing the reliability of a given probe and/or probe-set for molecular profiling and the methods of the present disclosure encompass any of these methods and combinations thereof.

III.B. Methods for Classifying IPF Subjects as Being at Risk for FVC Decline

The presently disclosed subject matter also provides in some embodiments methods for classifying subjects diagnosed with Idiopathic Pulmonary Fibrosis (IPF) as being at risk for a decline in lung Forced Vital Capacity (FVC). In some embodiments, the methods comprise determining a first expression level for one or more genes selected from the group consisting of *ALDH4A1*, *APTX*, *ATP6AP1L*, *CCNB1*, *CNR2*, *DNAJC17*, *DTWD1*, *FAM111B*, *GABRR1*, *GPR39*, *GYPA*, *HBB*, *HLA-DPB1*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MRPL35*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PLCL1*, *PNMA5*, *RAB3C*, *RBM43*, *RLBP1*, *SESN3*, *SLC25A37*, *SSU72P8*, *TP63*, *WDR17*, *ZNF252P*, and *ZNF582* in a first biological sample obtained from the subject diagnosed with IPF to establish a baseline expression level for the one or more genes; determining a second expression level for the one or more genes in a second biological sample obtained from the subject, wherein the first and second biological samples comprise peripheral blood mononuclear cells (PBMCs) and/or nucleic acids extracted from PBMCs; and comparing the first and second expression levels for the one or more genes to create an FVC-gene predictor score, wherein if the FVC-gene predictor score is greater than or equal to a pre-selected value, the patient is classified as being at risk for a decline in lung FVC within a pre-determined time period (e.g., 12 months) from the time that the first biological sample was obtained from the subject. With respect to the presently disclosed classifying methods, the methods comprise the same gene expression approaches disclosed herein above with respect to generating a prognostic signature for a given subject,

When expression levels for only one gene is being compared, the fold increase or fold decrease can be reported as a score. However, when expression levels for more than one gene are being compared, a synthesis of the various gene expression levels can be employed to generate an overall score. In some embodiments, a simple sum of the normalized fold increases (e.g., values ≥ 0) and normalized fold decreases (e.g., values ≤ 0) are employed to generate an overall score.

In some embodiments, the overall score is reported as a simple sum of the normalized fold increases and decreases, which in some embodiments can be referred to as

a “raw score”. In some embodiments, however, the overall score is reported as a weighted sum of the normalized fold increases and decreases. Values that can be employed for weighting can be pre-determined and can include, for example, using regression coefficients and assessed using area under the curve (AUC) analysis as described herein below. By way of example and not limitation, a score can be generated by multiplying normalized fold increase(s) and fold decrease(s) by a logistic LASSO regression coefficient derived from analyzing expression of any given gene(s) in normal controls and/or FVC-S patients, and summing the weighted values to produce an FVC-gene predictor score for a given subject.

Scores can vary based in some embodiments on the number of genes employed, in some embodiments on the assay technique employed, in some embodiments the time between the first and second sample isolations (e.g., between an initial isolation and an isolate 4 months later), and in some embodiments on a pre-selected minimum sensitivity. In some embodiments, a raw score of at least 5, 6, 7, or 8 can be indicative of a subject being in FVC decline when all 25 of *APTX*, *ATP6AP1L*, *CNR2*, *FAM111B*, *GABRR1*, *GPR39*, *GYP A*, *HBB*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PLCL1*, *PNMA5*, *RAB3C*, *RBM43*, *RLBP1*, *SSU72P8*, *TP63*, and *ZNF252P* are employed. In some embodiments, a score of at least 7.5 provides a 75% sensitivity when these 25 genes are employed and a microassay technique is employed for gene expression analysis. It is noted, however, that the technique employed for assaying gene expression changes and the time between first and second sample isolations can affect the values of the fold increases and decreases. As such, in some embodiments the same technique is employed for assaying gene expression at all times for both the control subjects and for the test subjects in order to minimize cross-testing variability, and the time between first and second sample isolations is fixed at four months. In some embodiments, when RNA-seq is employed to determine gene expression levels for the genes *APTX*, *CNR2*, *GABRR1*, *GPR39*, *GYP A*, *HBB*, *ITLN1*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PNMA5*, *RLBP1*, and *SSU72P8*, a score of -1.73 provides 75% sensitivity to the identification of FVC-D subjects.

III.C. Methods for Identifying and Treating IPF Subjects at Risk for FVC Decline

In some embodiments of the presently disclosed subject matter, once a patient is classified as being in or at risk for FVC-D, that patient can be identified as being appropriate for treatment, whereas if a patient is classified as not being in or at risk for FVC-D (i.e., is

classified as FVC-S), that patient can be identified as being appropriate for further monitoring but not treatment.

Accordingly, in some embodiments the presently disclosed subject matter relates to methods for identifying and treating IPF subjects at risk for or experience a decline in lung FVC. In some embodiments, the methods comprise determining a first expression level for one or more genes selected from the group consisting of *ALDH4A1*, *APTX*, *ATP6AP1L*, *CCNB1*, *CNR2*, *DNAJC17*, *DTWD1*, *FAM111B*, *GABRR1*, *GPR39*, *GYPA*, *HBB*, *HLA-DPB1*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MRPL35*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PLCL1*, *PNMA5*, *RAB3C*, *RBM43*, *RLBP1*, *SESN3*, *SLC25A37*, *SSU72P8*, *TP63*, *WDR17*, *ZNF252P*, and *ZNF582* in a first biological sample obtained from the subject diagnosed with IPF to establish a baseline expression level for the one or more genes; determining a second expression level for the one or more genes in a second biological sample obtained from the subject, wherein the first and second biological samples comprise peripheral blood mononuclear cells (PBMCs) and/or nucleic acids extracted from PBMCs; comparing the first and second expression levels for the one or more genes to create an FVC-gene predictor score; and if the FVC-gene predictor score is greater than or equal to a pre-selected value, treating the subject with a treatment selected from the group consisting of lung transplantation and a drug therapy. Appropriate treatments for patients in need of treatment include in some embodiments administering to the subject a pharmaceutical composition comprising pirfenidone, nintedanib, or a combination thereof in an amount and via a route of administration effective to delay or prevent the development of FVC decline in the subject. See U.S. Patent Nos. 3,974,281; 6,762,180; 8,592,462; 9,884,802; 10,028,966; and 10,105,365, each of which is incorporated herein by reference in its entirety. See also U.S. Patent Application Publication Nos. 2018/0064695, 2018/0169084; 2019/0030012; and 2019/0282565, each of which is incorporated herein by reference in its entirety.

As with other methods disclosed herein, in some embodiments the method comprise determining first and second expression levels for a set of genes selected from the group consisting of (a) *APTX*, *CNR2*, *GYPA*, *ITLN1*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PLA2G4A*, and *PNMA5*; (b) *APTX*, *ATP6AP1L*, *ITLN1*, *LINC00319*, *MAZ*, *MSR1*, *NT5E*, *PCDHB15*, *RAB3C*, *SSU72P8*, and *TP62*; (c) *APTX*, *CNR2*, *GABRR1*, *GPR39*, *GYPA*, *HBB*, *ITLN1*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PNMA5*, *RLBP1*, and *SSU72P8*; and (d) *APTX*, *ATP6AP1L*, *CNR2*, *FAM111B*, *GABRR1*, *GPR39*, *GYPA*, *HBB*, *IGLC1*, *ITLN1*,

LINC00319, MAZ, MSR1, NT5E, PAWR, PCDHB15, PLA2G4A, PLCL1, PNMA5, RAB3C, RBM43, RLBPI, SSU72P8, TP63, and ZNF252P.

The presently disclosed subject matter is also directed to methods of administering the compounds of the presently disclosed subject matter to a subject.

Pharmaceutical compositions comprising the present compounds are administered to a subject in need thereof by any number of routes including, but not limited to, topical, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal approaches.

In accordance with one embodiment, a method for treating a subject in need of such treatment is provided. The method comprises administering a pharmaceutical composition comprising at least one composition of the presently disclosed subject matter to a subject in need thereof. Compositions provided by the methods of the presently disclosed subject matter can be administered with known compounds or other medications as well.

The pharmaceutical compositions useful for practicing the presently disclosed subject matter may be administered to deliver a dose of between 1 ng/kg/day and 100 mg/kg/day.

The presently disclosed subject matter encompasses the preparation and use of pharmaceutical compositions comprising a compound useful for treatment of the diseases and disorders disclosed herein as an active ingredient. Such a pharmaceutical composition may consist of the active ingredient alone, in a form suitable for administration to a subject, or the pharmaceutical composition may comprise the active ingredient and one or more pharmaceutically acceptable carriers, one or more additional ingredients, or some combination of these. The active ingredient may be present in the pharmaceutical composition in the form of a physiologically acceptable ester or salt, such as in combination with a physiologically acceptable cation or anion, as is well known in the art.

As used herein, the term “physiologically acceptable” ester or salt means an ester or salt form of the active ingredient which is compatible with any other ingredients of the pharmaceutical composition, which is not deleterious to the subject to which the composition is to be administered.

The compositions of the presently disclosed subject matter may comprise at least one active polypeptide, one or more acceptable carriers, and optionally other polypeptides or therapeutic agents.

For in vivo applications, the compositions of the presently disclosed subject matter may comprise a pharmaceutically acceptable salt. Suitable acids which are capable of forming such salts with the compounds of the presently disclosed subject matter include inorganic acids such as hydrochloric acid, hydrobromic acid, perchloric acid, nitric acid, thiocyanic acid, sulfuric acid, phosphoric acid and the like; and organic acids such as formic acid, acetic acid, propionic acid, glycolic acid, lactic acid, anthranilic acid, cinnamic acid, naphthalene sulfonic acid, sulfanilic acid and the like.

Pharmaceutically acceptable carriers include physiologically tolerable or acceptable diluents, excipients, solvents, or adjuvants. The compositions are in some embodiments sterile and nonpyrogenic. Examples of suitable carriers include, but are not limited to, water, normal saline, dextrose, mannitol, lactose or other sugars, lecithin, albumin, sodium glutamate, cysteine hydrochloride, ethanol, polyols (propylene glycol, polyethylene glycol, glycerol, and the like), vegetable oils (such as olive oil), injectable organic esters such as ethyl oleate, ethoxylated isosteraryl alcohols, polyoxyethylene sorbitol and sorbitan esters, microcrystalline cellulose, aluminum methahydroxide, bentonite, kaolin, agar-agar and tragacanth, or mixtures of these substances, and the like.

The pharmaceutical compositions may also contain minor amounts of nontoxic auxiliary pharmaceutical substances or excipients and/or additives, such as wetting agents, emulsifying agents, pH buffering agents, antibacterial and antifungal agents (such as parabens, chlorobutanol, phenol, sorbic acid, and the like). Suitable additives include, but are not limited to, physiologically biocompatible buffers (e.g., tromethamine hydrochloride), additions (e.g., 0.01 to 10 mole percent) of chelants (such as, for example, DTPA or DTPA-bisamide) or calcium chelate complexes (as for example calcium DTPA or CaNaDTPA-bisamide), or, optionally, additions (e.g., 1 to 50 mole percent) of calcium or sodium salts (for example, calcium chloride, calcium ascorbate, calcium gluconate or calcium lactate). If desired, absorption enhancing or delaying agents (such as liposomes, aluminum monostearate, or gelatin) may be used. The compositions can be prepared in conventional forms, either as liquid solutions or suspensions, solid forms suitable for solution or suspension in liquid prior to injection, or as emulsions. Pharmaceutical compositions according to the presently disclosed subject matter can be prepared in a manner fully within the skill of the art.

The compositions of the presently disclosed subject matter or pharmaceutical compositions comprising these compositions may be administered so that the compositions

may have a physiological effect. Administration may occur enterally or parenterally; for example, orally, rectally, intracisternally, intravaginally, intraperitoneally, locally (e.g., with powders, ointments or drops), or as a buccal or nasal spray or aerosol. Parenteral administration is an approach. Particular parenteral administration methods include intravascular administration (e.g., intravenous bolus injection, intravenous infusion, intra-arterial bolus injection, intra-arterial infusion and catheter instillation into the vasculature), peri- and intra-target tissue injection, subcutaneous injection or deposition including subcutaneous infusion (such as by osmotic pumps), intramuscular injection, and direct application to the target area, for example by a catheter or other placement device.

Where the administration of the composition is by injection or direct application, the injection or direct application may be in a single dose or in multiple doses. Where the administration of the compound is by infusion, the infusion may be a single sustained dose over a prolonged period of time or multiple infusions.

The formulations of the pharmaceutical compositions described herein may be prepared by any method known or hereafter developed in the art of pharmacology. In general, such preparatory methods include the step of bringing the active ingredient into association with a carrier or one or more other accessory ingredients, and then, if necessary or desirable, shaping or packaging the product into a desired single- or multi-dose unit.

It will be understood by the skilled artisan that such pharmaceutical compositions are generally suitable for administration to animals of all sorts. Subjects to which administration of the pharmaceutical compositions of the presently disclosed subject matter is contemplated include, but are not limited to, humans and other primates, mammals including commercially relevant mammals such as cattle, pigs, horses, sheep, cats, and dogs, birds including commercially relevant birds such as chickens, ducks, geese, and turkeys.

A pharmaceutical composition of the presently disclosed subject matter may be prepared, packaged, or sold in bulk, as a single unit dose, or as a plurality of single unit doses. As used herein, a "unit dose" is a discrete amount of the pharmaceutical composition comprising a predetermined amount of the active ingredient. The amount of the active ingredient is generally equal to the dosage of the active ingredient which would be administered to a subject or a convenient fraction of such a dosage such as, for example, one-half or one-third of such a dosage.

The relative amounts of the active ingredient, the pharmaceutically acceptable carrier, and any additional ingredients in a pharmaceutical composition of the presently

disclosed subject matter will vary, depending upon the identity, size, and condition of the subject treated and further depending upon the route by which the composition is to be administered. By way of example, the composition may comprise between 0.1% and 100% (w/w) active ingredient.

In addition to the active ingredient, a pharmaceutical composition of the presently disclosed subject matter may further comprise one or more additional pharmaceutically active agents. Particularly contemplated additional agents include anti-emetics and scavengers such as cyanide and cyanate scavengers.

Controlled- or sustained-release formulations of a pharmaceutical composition of the presently disclosed subject matter may be made using conventional technology.

As used herein, “additional ingredients” include, but are not limited to, one or more of the following: excipients; surface active agents; dispersing agents; inert diluents; granulating and disintegrating agents; binding agents; lubricating agents; sweetening agents; flavoring agents; coloring agents; preservatives; physiologically degradable compositions such as gelatin; aqueous vehicles and solvents; oily vehicles and solvents; suspending agents; dispersing or wetting agents; emulsifying agents, demulcents; buffers; salts; thickening agents; fillers; emulsifying agents; antioxidants; antibiotics; antifungal agents; stabilizing agents; and pharmaceutically acceptable polymeric or hydrophobic materials. Other “additional ingredients” which may be included in the pharmaceutical compositions of the presently disclosed subject matter are known in the art and described, for example in Gennaro, 1990 and/or Gennaro, 2003, each of which is incorporated herein by reference.

Typically, dosages of the compound of the presently disclosed subject matter which may be administered to an animal, in some embodiments a human, range in amount from 1 μ g to about 100 g per kilogram of body weight of the animal. While the precise dosage administered will vary depending upon any number of factors, including but not limited to, the type of animal and type of disease state being treated, the age of the animal and the route of administration. In some embodiments, the dosage of the compound will vary from about 1 mg to about 10 g per kilogram of body weight of the animal. In another aspect, the dosage will vary from about 10 mg to about 1 g per kilogram of body weight of the animal.

The compositions may be administered to an animal as frequently as several times daily, or it may be administered less frequently, such as once a day, once a week, once every two weeks, once a month, or even less frequently, such as once every several months or even once a year or less. The frequency of the dose will be readily apparent to the skilled

artisan and will depend upon any number of factors, such as, but not limited to, the type of cancer being diagnosed, the type and severity of the condition or disease being treated, the type and age of the animal, etc.

Suitable preparations include injectables, either as liquid solutions or suspensions, however, solid forms suitable for solution in, suspension in, liquid prior to injection, may also be prepared. The preparation may also be emulsified, or the compositions encapsulated in liposomes. The active ingredients are often mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient. Suitable excipients are, for example, water saline, dextrose, glycerol, ethanol, or the like and combinations thereof. In addition, if desired, the preparation may also include minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents, and/or adjuvants.

Various aspects and embodiments of the presently disclosed subject matter are described in further detail below.

The formulations of the pharmaceutical compositions described herein may be prepared by any method known or hereafter developed in the art of pharmacology. In general, such preparatory methods include the step of bringing the active ingredient into association with a carrier or one or more other accessory ingredients, and then, if necessary or desirable, shaping or packaging the product into a desired single- or multi-dose unit.

The compounds of the presently disclosed subject matter may be administered to, for example, a cell, a tissue, or a subject by any of several methods described herein and by others which are known to those of skill in the art.

The relative amounts of the active ingredient, the pharmaceutically acceptable carrier, and any additional ingredients in a pharmaceutical composition of the presently disclosed subject matter will vary, depending upon the identity, sex, age, size, and condition of the subject treated and further depending upon the route by which the composition is to be administered.

In addition to the active ingredient, a composition of the presently disclosed subject matter may further comprise one or more additional pharmaceutically active or therapeutic agents. Particularly contemplated additional agents include anti-emetics and scavengers such as cyanide and cyanate scavengers.

Controlled- or sustained-release formulations of a composition of the presently disclosed subject matter may be made using conventional technology.

As used herein, “additional ingredients” include, but are not limited to, one or more of the following: excipients; surface active agents; dispersing agents; inert diluents; granulating and disintegrating agents; binding agents; lubricating agents; sweetening agents; flavoring agents; coloring agents; preservatives; physiologically degradable compositions such as gelatin; aqueous vehicles and solvents; oily vehicles and solvents; suspending agents; dispersing or wetting agents; emulsifying agents, demulcents; buffers; salts; thickening agents; fillers; emulsifying agents; antioxidants; antibiotics; antifungal agents; stabilizing agents; and pharmaceutically acceptable polymeric or hydrophobic materials. Other “additional ingredients” which may be included in the pharmaceutical compositions of the presently disclosed subject matter are known in the art and described, for example in Gennaro, 1990 and/or Gennaro, 2003, each of which is incorporated herein by reference.

Other components such as preservatives, antioxidants, surfactants, absorption enhancers, viscosity enhancers or film forming polymers, bulking agents, diluents, coloring agents, flavoring agents, pH modifiers, sweeteners or taste-masking agents may also be incorporated into the composition. Suitable coloring agents include red, black, and yellow iron oxides and FD&C dyes such as FD&C Blue No. 2, FD&C Red No. 40, and the like. Suitable flavoring agents include mint, raspberry, licorice, orange, lemon, grapefruit, caramel, vanilla, cherry grape flavors, combinations thereof, and the like. Suitable pH modifiers include citric acid, tartaric acid, phosphoric acid, hydrochloric acid, maleic acid, sodium hydroxide, and the like. Suitable sweeteners include aspartame, acesulfame K, thaumatic, and the like. Suitable taste-masking agents include sodium bicarbonate, ion-exchange resins, cyclodextrin inclusion compounds, adsorbates, and the like.

The formulations of the pharmaceutical compositions described herein may be prepared by any method known or hereafter developed in the art of pharmacology. In general, such preparatory methods include the step of bringing the active ingredient into association with a carrier or one or more other accessory ingredients, and then, if necessary or desirable, shaping or packaging the product into a desired single- or multi-dose unit.

Although the descriptions of pharmaceutical compositions provided herein are principally directed to pharmaceutical compositions which are suitable for ethical administration to humans, it will be understood by the skilled artisan that such compositions are generally suitable for administration to animals of all sorts. Modification of pharmaceutical compositions suitable for administration to humans in order to render the compositions suitable for administration to various animals is well understood, and the

ordinarily skilled veterinary pharmacologist can design and perform such modification with merely ordinary, if any, experimentation. Subjects to which administration of the pharmaceutical compositions of the presently disclosed subject matter is contemplated include, but are not limited to, humans and other primates, mammals including commercially relevant mammals such as cattle, pigs, horses, sheep, cats, and dogs, and birds including commercially relevant birds such as chickens, ducks, geese, and turkeys.

The pharmaceutical compositions of the presently disclosed subject matter can be administered in any suitable formulation, by any suitable means, and by any suitable route of administration. Formulations suitable for topical administration include, but are not limited to, liquid or semi-liquid preparations such as liniments, lotions, oil in water or water in oil emulsions such as creams, ointments or pastes, and solutions or suspensions. Topically-administrable formulations may, for example, comprise from about 1% to about 10% (w/w) active ingredient, although the concentration of the active ingredient may be as high as the solubility limit of the active ingredient in the solvent. Formulations for topical administration may further comprise one or more of the additional ingredients described herein.

An alternative standard of care treatment for patients diagnosed with FVC-D and/or who are at risk for developing FVC-D within a pre-determined time period is lung transplantation. Thus, in some embodiments a patient classified and/or identified with FVC-D and/or who is at risk for developing FVC-D within a pre-determined time period (e.g., within 12 months) is an appropriate candidate for lung transplantation.

III.D. Methods for Monitoring the Progress of a Treatment

The basic techniques described herein can also be employed to monitor the progress of a treatment. As used herein, the phrase “progress of a treatment” refers to the ability of a treatment to reduce FVC decline over time, particularly with respect to reducing the rate at which FVC decline occurs in a patient.

As such, in some embodiments the presently disclosed subject matter relates to methods for monitoring the progress of a treatment in an IPF patient whose is experiencing a decline in lung Forced Vital Capacity (FVC) comprising determining a first expression level for one or more genes selected from the group consisting of *ALDH4A1*, *APTX*, *ATP6AP1L*, *CCNB1*, *CNR2*, *DNAJC17*, *DTWD1*, *FAM111B*, *GABRR1*, *GPR39*, *GYP A*, *HBB*, *HLA-DPB1*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MRPL35*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PLCL1*, *PNMA5*, *RAB3C*, *RBM43*, *RLBP1*, *SESN3*, *SLC25A37*,

SSU72P8, *TP63*, *WDR17*, *ZNF252P*, and *ZNF582* in a first biological sample obtained from the patient to establish a baseline expression level for the one or more genes; determining a second expression level for the one or more genes in a second biological sample obtained from the patient at a subsequent time point, wherein the first and second biological samples comprise peripheral blood mononuclear cells (PBMCs) and/or nucleic acids extracted from PBMCs; and comparing the first and second expression levels for the one or more genes, wherein the comparing step is indicative of the progress of the treatment in the patient.

As set forth herein above, an exemplary treatment for FVC-decline comprises administering to the patient an effective amount of pirfenidone, nintedanib, or a combination thereof. At a first time point (including but not limited to a time point at or before initiation of the treatment), a normalized expression level for each gene in first biological sample can be determined. Thereafter, at a subsequent timepoint of interest (e.g., one or more weeks or months subsequent to the initial timepoint), the same genes can again be assayed and a normalized expression level for each gene in the second, subsequent biological sample can be determined. In some embodiments, the first and second normalized expression level for each gene assays are compared to generate a fold-increase and/or a fold-decrease in the second biological sample relative to the first biological sample for each gene. As before, the comparing can comprise summing each fold-increase and/or fold-decrease to produce an FVC-gene predictor score for the patient, wherein the FVC-gene predictor score produced can be a raw or a weighted score.

Also as before, the set of genes for which first and second expression levels are determined can be selected from the group consisting of: (a) *APTX*, *CNR2*, *GYP A*, *ITLN1*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PLA2G4A*, and *PNMA5*; (b) *APTX*, *ATP6AP1L*, *ITLN1*, *LINC00319*, *MAZ*, *MSR1*, *NT5E*, *PCDHB15*, *RAB3C*, *SSU72P8*, and *TP62*; (c) *APTX*, *CNR2*, *GABRR1*, *GPR39*, *GYP A*, *HBB*, *ITLN1*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PNMA5*, *RLBP1*, and *SSU72P8*; and (d) *APTX*, *ATP6AP1L*, *CNR2*, *FAM111B*, *GABRR1*, *GPR39*, *GYP A*, *HBB*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PLCL1*, *PNMA5*, *RAB3C*, *RBM43*, *RLBP1*, *SSU72P8*, *TP63*, and *ZNF252P*. In some embodiments, first and second expression levels for each of *APTX*, *ATP6AP1L*, *CNR2*, *FAM111B*, *GABRR1*, *GPR39*, *GYP A*, *HBB*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PLCL1*, *PNMA5*, *RAB3C*, *RBM43*, *RLBP1*, *SSU72P8*, *TP63*, and *ZNF252P* are compared.

In some embodiments, the second biological sample is obtained from the patient at a time subsequent to when the first biological sample was obtained from the patient selected from the group consisting of about 1 week, about 2 weeks, about 4 weeks, about 6 weeks, about 2 months, about 3 months, about 4 months, about 5 months, about 6 months, or longer than six months.

In some embodiments, the presently disclosed monitoring method can further comprise determining a one or more subsequent expression levels for one or more genes selected from the group consisting of *ALDH4A1*, *APTX*, *ATP6AP1L*, *CCNB1*, *CNR2*, *DNAJC17*, *DTWD1*, *FAM111B*, *GABRR1*, *GPR39*, *GYPA*, *HBB*, *HLA-DPB1*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MRPL35*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PLCL1*, *PNMA5*, *RAB3C*, *RBM43*, *RLBP1*, *SESN3*, *SLC25A37*, *SSU72P8*, *TP63*, *WDR17*, *ZNF252P*, and *ZNF582* in one or more subsequently isolated biological samples obtained from the patient; and (e) comparing the first, second, and one or more subsequent expression levels for the one or more genes, wherein the comparing step is indicative of the progress of the treatment in the patient.

In accordance with the presently disclosed subject matter, as described above or as discussed in the EXAMPLES below, there can be employed conventional chemical, cellular, histochemical, biochemical, molecular biology, microbiology, recombinant DNA, and clinical techniques which are known to those of skill in the art. Such techniques are explained fully in the literature. See for example, Sambrook et al., 1989; Glover, 1985; Gait, 1984; Harlow & Lane, 1988; Roe et al., 1996; and Ausubel et al., 1995.

The presently disclosed subject matter may be embodied in other specific forms without departing from the spirit or essential attributes thereof. The presently disclosed subject matter encompasses all combinations of the different aspects of the presently disclosed subject matter noted herein. It is understood that any and all embodiments of the presently disclosed subject matter may be taken in conjunction with any other embodiment or embodiments to describe additional representative embodiments. It is also to be understood that each individual element of the disclosed embodiments is intended to be taken individually as its own independent representative embodiment. Furthermore, any element of an embodiment is meant to be combined with any and all other elements from any embodiment to describe an additional embodiment.

Typically, dosages of the compounds of the presently disclosed subject matter which may be administered to an animal, in some embodiments a human, range in amount from

about 1.0 µg to about 100 g per kilogram of body weight of the animal. The precise dosage administered will vary depending upon any number of factors, including but not limited to, the type of animal and type of disease state being treated, the age of the animal and the route of administration. In some embodiments, the dosage of the compound will vary from about 1 mg to about 10 g per kilogram of body weight of the animal. In some embodiments, the dosage will vary from about 10 mg to about 1 g per kilogram of body weight of the animal.

The compounds may be administered to a subject as frequently as several times daily, or it may be administered less frequently, such as once a day, once a week, once every two weeks, once a month, or even less frequently, such as once every several months or even once a year or less. The frequency of the dose will be readily apparent to the skilled artisan and will depend upon any number of factors, such as, but not limited to, the type and severity of the disease being treated, the type and age of the animal, etc.

EXAMPLES

The presently disclosed subject matter will be now be described more fully hereinafter with reference to the accompanying EXAMPLES, in which representative embodiments of the presently disclosed subject matter are shown. The presently disclosed subject matter can, however, be embodied in different forms and should not be construed as limited to the embodiments set forth herein. Rather, these embodiments are provided so that this disclosure will be thorough and complete, and will fully convey the scope of the presently disclosed subject matter to those skilled in the art.

Materials and Methods

IPF Cohorts Generally. The training cohort was comprised of patients participating in the prospective COMET study (Huang et al., 2017; NCT01071707). Validation cohorts included prospectively enrolled patients at the University of Chicago (UChicago); University of Pittsburgh Medical Center (UPMC; Herazo-Maya et al., 2013); and Imperial College London (Imperial; Herazo-Maya et al., 2017). All patients were diagnosed with IPF according to international guidelines (American Thoracic Society & European Respiratory Society, 2002; Raghu et al., 2011). Patients in each cohort were stratified according to the presence of progressive disease, defined as $\geq 10\%$ relative decline in FVC over the study timeframe. Additional cohort-specific detailed are provided in the online supplement.

COMET Training Cohort. Subjects included in this analysis were participants in COMET-IPF (Correlating Outcomes with biochemical Markers to Estimate Time-progression in Idiopathic Pulmonary Fibrosis), a prospective, observational study

correlating biomarkers with disease progression (NCT01071707; Naik et al., 2012). This multicenter investigation recruited subjects at nine clinical centers in the US. Inclusion criteria required diagnosis of IPF was confirmed using a multidisciplinary diagnostic approach per international guidelines (Raghu et al., 2011) using expertise from clinicians, radiologists, and pathologists at the local, enrolling clinical center (Flaherty et al., 2004; Flaherty et al., 2007) and age 35-80 years. Subjects were excluded if the diagnosis of IPF was >4 years prior to screening or if there was a diagnosis of collagen-vascular disorder, FEV1/FVC <0.60, evidence of active infection at screening, or comorbid conditions other than IPF likely to result in death within one year. Subjects underwent protocol-directed visits every 4 months after the baseline (0 visit) for a minimum of 1 year, establishing four transcriptome sampling timepoints with PFTs and blood draws performed every 4 months. Registry patients with peripheral blood mononuclear cells (PBMC) gene expression (GE) sampling over at least two time points were included in training (1-4 month) and each subset cohort (i.e., 0-8 month; 0-12 month; 4-8 month; 4-12 month). Forced vital capacity (FVC) and diffusion capacity for carbon monoxide (DLCO) were obtained per ATS guidelines (Macintyre et al., 2005; Miller et al., 2005a; Miller et al., 2005b). Subjects in COMET experiencing a relative reduction < 10% or $\geq 10\%$ in FVC% predicted from baseline visit to follow-up visit at month 12 were defined as FVC-S (stable) or FVC-D (decline), respectively. DLCO-S (stable) or DLCO-D (decline) were defined as <15% or $\geq 15\%$ of relative reduction in DLCO% predicted from the baseline visit to follow-up at month 12, respectively. Informed consent was obtained from all participants. The study protocol was reviewed by the institutional review board of each participating center.

UChicago Validation Cohort. Study populations were collected from the University of Chicago Medical Center and was approved by the institutional review board and informed consent was provided by all study subjects. All patients with IPF met American Thoracic Society/European Respiratory Society (ATS/ERS) diagnosis criteria (American Thoracic Society & European Respiratory Society, 2002). Demographic information, clinical characteristics, and pulmonary function tests were collected from all patients with IPF. Spirometry testing, including forced vital capacity percent predicted (FVC% predicted), diffusion capacity for carbon monoxide percent predicted (DLCO% predicted) as well as lung volumes by plethysmography were obtained per ATS guidelines (Macintyre et al., 2005; Miller et al., 2005a; Miller et al., 2005b). The prognosis of IPF subjects was

dichotomously categorized as FVC stable (FVC-S) or FVC decline (FVC-D) defined by $<$ or $\geq 10\%$ reduction in FVC% predicted from the baseline to over 2 years of follow-up.

UPMC validation cohort. Patients were recruited from the University of Pittsburgh. IPF diagnosis was established by a multidisciplinary group at each institution with the American Thoracic Society/European Respiratory Society criteria (American Thoracic Society & European Respiratory Society, 2002) and was consistent with recent guidelines (Raghu et al., 2011). Patients were excluded in the study if they had evidence of autoimmune syndromes, malignancies, infections, drugs, or occupational exposures known to cause lung fibrosis. The studies were approved by the institutional review boards at the two institutions, and informed consent was obtained from all patients. Demographic and clinical information were collected in all patients at the time of blood draw. Spirometric data and diffusion capacity of the lung for carbon monoxide (DLCO) were obtained within 3 months of blood draw, with the exception of four IPF patients of the replication cohort who did not have DLCO values available within this time range. The prognosis of IPF subjects was dichotomously categorized as FVC stable (FVC-S) or FVC decline (FVC-D) defined by $<$ or $\geq 10\%$ relative reduction in FVC% of predicted from the baseline to about 12 month follow-up.

Imperial Validation Cohort. Patients were prospectively recruited from the Interstitial Lung Disease Unit at the Royal Brompton Hospital, London, United Kingdom, between November 2010 and January 2013. Diagnoses of IPF were made according to international guidelines (Raghu et al., 2011) after multidisciplinary team discussion. Subjects were excluded if they had a history of self-reported upper or lower respiratory tract infection, antibiotic use in the prior 3 months, acute IPF exacerbation, or other respiratory disorders. Written informed consent was obtained from all subjects, and the study was approved by the local research ethics committee (reference numbers 10/H0720/12 and 12/LO/1034). At baseline and at each subsequent visit, pulmonary function test was performed and peripheral blood were collected into PAXgene RNA tubes (PreAnalytiX, Hombrechtikon, Switzerland). at baseline and at 6 and 12 months. The prognosis of IPF subjects was dichotomously categorized as FVC stable (FVC-S) or FVC decline (FVC-D) defined by $<$ or $\geq 10\%$ relative reduction in FVC% of predicted from the baseline to about 12 month follow-up.

Gene Expression Data. Information regarding gene expression (GE) assays, raw data processing and normalization, pathway analyses, and sample classification were as follows.

COMET Cohort. PBMC sample collection, RNA isolation, microarray hybridization, and data processing. Peripheral blood mononuclear cells (PBMCs) from IPF patients were isolated from whole blood collected in Lavender top tubes containing EDTA by Ficoll-Paque Plus (GE Healthcare Life Science, Pittsburgh, Pennsylvania, United States of America) as described in Eppendorf Application Note No. 372 dated June 2016 (available from the Eppendorf website) and lysed with TRIzol reagent (Thermo Fisher Sci., Waltham, Massachusetts, United States of America) for RNA extraction following manufacturer's protocol. Sodium acetate/ethanol was used to re-precipitate RNA to increase the purity prior to Affymetrix PRIMEVIEW™ brand array assay (Affymetrix, Santa Clara, CA) according to manufacturer's manual (available from the Affymetric website). RNA quality and integrity were confirmed by Nanodrop (A_{260}/A_{280} ratios between 1.7 and 2.2) and Bio-Analyzer mini-gel assay, respectively. One hundred fifty ng RNA per sample was reverse transcribed to single stranded cDNA, and then amplified to cRNA using Affymetrix GeneChip WT cDNA Synthesis Kit. Qualities and yields (exceeding 25 µg/ml and 1000 µg/ml, respectively) of cRNA after the first and second amplifications were all satisfactory prior to hybridization and scanning. The Affymetrix microarray raw data in ".cel" format were processed using R/Bioconductor package "affy" (Gautier et al., 2004). Background correction and gene expression intensities were summarized and normalized using "rma" algorithm (Irizarry et al., 2003). The complete data sets are available in the Gene Expression Omnibus database under platform No. GSE132607 (see the website of the National Center for Biotechnology Information of the United States National Institutes of Health; Accession No. GSE132607).

UChicago Cohort. PBMC sample collection, RNA isolation, RNA-Seq library preparation and sequencing. PBMC samples were obtained by density centrifugation. RNA was extracted with TRIzol (Invitrogen) and was re-precipitated by sodium acetate/ethanol. RNA quality and integrity were confirmed by Nanodrop (A_{260}/A_{280} ratios between 1.7 and 2.2) and Bio-Analyzer mini-gel assay (Agilent, Santa Clara, California, United States of America). All RNA samples displayed a RNA Integrity Number (RIN) > 7 were proceeded to cDNA library preparation at the Genomics Core Facility of the University of Chicago. Total RNA in the amount of 1 µg per sample was depleted of ribosomal RNA using the Ribo-Zero kit (Epicentre, Madison, Wisconsin, United States of America). The directional (first strand) cDNA libraries were prepared following the guide of TruSeq Stranded Total RNA Sample Preparation kit. RNA was fragmented at 94°C for 6 minutes, followed by the

first strand cDNA generation. Deoxy-UTP was incorporated in second strand synthesis in order to effectively quench the second strand during PCR amplification. After adenylation of the 3' end and ligation of adapters, fragments were selected and enriched with 10 cycles of PCR amplification. Clusters were generated by bridge amplification within paired-end flow cells using Illumina HiSeq PE Cluster Kit v4 cBot according to manufacturer's instructions (Illumina, San Diego, California, United States of America). The clusters on flow cells were then sequenced on the Illumina HiSeq4000 using HiSeq SBS Kit. A total of 1 Tbase reads were generated for cDNA libraries prepared from 54 samples using high output mode of 100bp paired-end (PE) sequencing. Around 94% sequences passed quality checked ($> Q30$), yielding about 87M passing filter clusters per sample, and 4.7G clusters in total. Raw sequencing data in fastq format were processed using RNA-seq aligner STAR v2 (Dobin et al., 2013). GenCode v24 was used for transcriptome annotation. The abundance of transcripts was summarized into CPM (Counts per Million mapped reads). Genes with a value of CPM > 0.5 in at least two samples were included for downstream analysis. The filtered raw read data were transformed to $\log_2$ -counts per million (logCPM) and normalized with the associated precision weights using "voom" (Law et al., 2014) and "TMM" (Robinson & Oshlack, 2010) normalization implemented in R/Bioconductor packages, including "limma" (Smyth, 2004) and "edgeR" (Robinson et al., 2010). Given the different technologies of gene expression assay used in COMET (Affymetrix PrimeView) and the three independent validation cohorts (Agilent 4x44K, Affymetrix Human Gene 1.1 ST and RNA-seq), the Gene ID was matched across different platforms.

UPMC Cohort. PBMC sample collection, RNA isolation, microarray hybridization, and data processing. Peripheral blood was collected in a cell preparation tube, followed by centrifugation to isolate PBMCs. These cells were suspended in QIAzol (Qiagen) and stored at -80°C . Total RNA was extracted and purified using the miRNeasy Mini Kit (Qiagen) and QIAcube device (Qiagen), following the manufacturer's protocols. After extraction, total RNA yield and quality were evaluated using NanoDrop at 260 nm and the 2100 Bioanalyzer (Agilent Technologies). Labeling reactions were performed using Agilent Quick Amp labeling kit, one-color (Agilent Technologies). Briefly, an initial cDNA strand was synthesized using 400 ng of total RNA and a T7-oligo(dT) primer containing a phage T7 RNA Polymerase promoter sequence at its 5'-end. This cDNA was then used as a template to generate Cy3-labeled cRNA by a reverse transcriptase enzyme. The cRNA was fragmented, hybridized to Whole Human Genome Oligo Microarray, 4 x 44K (G4112F,

Agilent Technologies), and scanned using an Agilent Microarray Scanner. For array readout, Agilent Feature Extraction software version 10.7 was used. To normalize the gProcessed signal, cyclic-LOESS was performed using the bioconductor package as described previously (Ballman et al., 2004). The average of the gene expression signal was used in the case of replicated probes for the same gene with different expression values. The complete data sets are available in the Gene Expression Omnibus database under Accession No. GSE28221 via the website of the National Center for Biotechnology Information of the United States National Institutes of Health.

Imperial cohort. PBMC sample collection, RNA isolation, microarray hybridization, and data processing. Whole blood was collected using PAXgene blood RNA tubes (PreAnalytiX) and stored in -80°C. Total RNA extraction was performed using the PAXgene Blood RNA Kit, following the manufacturer's protocol. Total RNA was quantified using the NanoDrop ND 1000 UV-Vis spectrophotometer (Thermo Scientific, Wilmington, DE), and the quality and integrity were assessed using the 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California, United States of America) by ratio comparison of the 18S and 28S rRNA bands. Thirty nanograms of each RNA sample was used to synthesize double-stranded complementary DNA (dscDNA) using the Ovation Pico WTA System V2 Kit (NuGEN, San Carlos, California, United States of America). Exogenous poly(A)-positive control subjects were added to monitor the efficiency of the synthesis of the dscDNA and target-labeling process. The Encore Biotin Module Kit (NuGEN) was used to fragment 2.8 µg of the purified cDNA template, which was then hybridized, washed, and scanned on the GeneTitan system (Affymetrix, Santa Clara, CA) using Human Gene 1.1 ST 16- or 24-sample array plates (Affymetrix). The complete data sets are available in the Gene Expression Omnibus database under Accession No. GSE93606 via the website of the National Center for Biotechnology Information of the United States National Institutes of Health.

Gene ID was matched across cohorts to account for differing GE assay platforms.

FVC-gene Predictor Training and Validation. Gene expression changes (Δ GE) between baseline and 4 month visit were compared between stable and progressive groups in the COMET training cohort using empirical Bayesian moderated t-test implemented in R/Bioconductor package "limma" (Smyth, 2004). P-values were adjusted for multiple comparisons using the Benjamini-Hochberg method (Benjamini & Hochberg, 1995). The R package 'glmet' (Friedman et al., 2010; Simon et al., 2011; Tibshirani et al., 2012) was then

used to perform Logistic Least Absolute Shrinkage and Selection Operator (LASSO) to enhance the prediction accuracy via variable selection and regularization. Ten-fold Cross-Validation was performed in conjunction with logistic LASSO regression to evaluate correct classification rate.

Genes predictive of FVC decline by this approach were used to generate an FVC-gene predictor score, defined as the sum of each Δ GE value multiplied by the corresponding logistic LASSO regression coefficient. Receiver operator analysis was performed using R-CRAN package “pROC” (Robin et al., 2011) and “OptimalCutpoints” (López-Ratón et al., 2014) to identify the optimal threshold for cohort stratification. Scores above that threshold were considered to have a positive FVC-gene predictor. The FVC-gene predictor score was then calculated for patients in each validation cohort using the subset of overlapping genes from each platform weighted by the cohort-specific cross-validation coefficient to identify those with a positive FVC-gene predictor. FVC-gene predictor test performance characteristics were then assessed in each cohort.

Comparative Analysis between FVC-gene Predictor and Cross-sectional Biomarkers. The test performance for predicting FVC decline was then compared between the FVC-gene predictor and prior plasma biomarkers of IPF mortality, including circulating plasma MMP7, periostin (POSTN), and CCL18. Additional information about this analysis is detailed in the online supplement.

Comparison of Cross-sectional and Longitudinal Gene Expression Modeling. The Coefficient of Variation (CoV) was compared between two different approaches utilizing gene expression data to determine impact on development and performance of an FVC-gene predictor. The CoV was calculated for each gene by dividing the gene-specific standard deviation by the mean. Intra-subject CoV for each gene was then computed using the root mean square method, where d is the difference between two paired measurements and m is the mean of paired measurements (Hyslop & White, 2009).

Sample Classification. Sample and gene clustering based on a priori selected genes was performed using dChip software (Li, 2008) or R/Bioconductor package ‘ComplexHeatmap’ (Gu et al., 2016). Principle Component Analysis (PCA) was performed using R package ‘FactoMine’ (Husson et al., 2010).

Pathway Analyses. Gene expression changes (Δ GE) between baseline and 4 month visit were compared between FVC-D and FVC-S groups. Gene set enrichment analysis (GSEA; Subramanian et al., 2005) of Δ GE was conducted at whole transcriptome level to

identify significant canonical pathways with the criterion of false discover rate (FDR) <10%. Positive enrichment scores (ES) represented longitudinal decrease of gene expressions from baseline to 4 month follow-up, while negative ES represented longitudinal increase in gene expressions in COMET training cohort.

Genes constituting the FVC-gene predictor were analyzed using R package “GOSim” (Alexa et al., 2006) with the criterion of q-value (Benjamini-Yekutieli adjusted p-value) <0.01 for biological process enrichment. Alternatively, ToppFun web application (Kaimal et al., 2010) was used for functional enrichment analysis of FVC-gene predictor.

Peripheral Plasma Proteomics. The COMET baseline plasma samples had previously been subjected to a proteomic aptamer-based assay (O’Dwyer et al., 2017). Levels of MMP7 (Rosas et al., 2008), CCL18 (PARC; Prasse et al., 2009; Hoffmann-Vold et al., 2016), and Periostin (POSTN; Naik et al., 2012; Tajiri et al., 2015; Ohta et al., 2017) and their predictive/prognostic performance of absolute change in FVC% predicted were extracted (Neighbors et al., 2018) and conducted as comparators for ROC/AUC analyses of the FVC-gene predictor. Peripheral blood was collected in EDTA-containing vacutainers at study centers and samples were prepared as described before (O’Dwyer et al., 2017). The SOMASCAN® proteomic assay has been described. In brief, each of the listed proteins is measured using a modified aptamer reagent and measured quantitatively in relative fluorescence units (RFU’s) using a custom Agilent hybridization chip. Normalization and inter-run calibration were performed according to SOMASCAN® v3 assay data quality-control procedures as defined in the SomaLogic good laboratory practice quality system. A complete list of SOMASCAN® analytes can be found online via the website of Somalogic).

EXAMPLE 1

Study Cohort Characteristics

Demographics, median follow-up time, median transcriptomic sampling timepoints and interquartile range (IQR), PFT, and outcomes for the COMET and validation cohorts are shown in Table 2. Twenty-two percent (16/74) of patients experienced FVC decline in the COMET training cohort and ranged from 30-63% in the validation cohorts, with the Imperial cohort having the highest prevalence of progressors. No significant differences were noted in the COMET training cohort with regard to demographics or lung function when stratifying by FVC decline (Table 3). The mean time from 2nd blood draw to the PFT follow-up was 12.1 months in UChicago, 6.5 months in UPMC, and 10.7 months in Imperial.

Table 2
Clinical Demographics in IPF Cohorts

Clinical Demographics	COMET Cohort				Independent Validation		
	0-4 mo	0-8 mo	4-8 mo	0-12 mo	UChicago	UPMC	Imperial
Sample Size	74	67	67	59	27	35	24
GE sampling duration Median (IRQ)	4 mo (n/a)	8 mo (n/a)	4 mo (n/a)	12 mo (n/a)	16.6 mo (9.3-25.9)	6 mo (2-8)	6 mo (n/a)
FVC follow-up Median (IRQ)	12 mo (n/a)	12 mo (n/a)	12 mo (n/a)	12 mo (n/a)	28 mo (20.2-35.8)	13 mo (10-16)	12 mo (7.8-16.2)
Sex (M/F)	52/22	47/20	47/20	44/15	19/8	21/4	18/6
Caucasian (%)	94.6	94.0	94.0	94.9	85.2	92.3	94.3
Smokers (%)	66.2	67.2	70.1	67.8	48.1	73.1	57.1
Age (average)	66.6	67.0	67.5	65.9	66.7	68.3	66.8
FVC (Stable/Progressor)	58/16	54/13	54/13	45/14	19/8	23/12	9/15
DLCO (Stable/Progressor)	48/26	42/25	43/24	40/19	9/14	15/20	5/17

Table 3
Stratification of COMET Training Cohort by FVC Status

FVC Status*	Sex (M/F)	Race [#] (C/O)	Smoker (Y/N)	Age (Mean)	Baseline ¹ FVC_pp	Baseline ² DLCO_pp
Stable	41/17	55/3	38/20	67.4	71.6	46.4
Progressive	11/5	15/1	10/6	63.7	62.6	42.7

*FVC status was defined as relative decline of FVC % of predicted over 12 months from baseline. [#]Race (C/O): Caucasian or Others; ¹FVC_pp: FVC percent of predicted; ²DLCO_pp: DLCO percent of predicted.

EXAMPLE 2

FVC-gene Predictor Training

A flowchart of study design and data analysis processes is illustrated in Figures 1A and 1B. An empirical Bayes moderated t-test identified 3906 probe-sets at FDR<0.05

predictive of FVC decline using Δ GE training data (Figure 1A-2). LASSO regression further reduced this to 39 genes identified, which correctly classified 88% ($n = 65/74$) of patients in the training set. 25 out of the 39 genes displaying $\geq 50\%$ cross-validation support were prioritized (Figure 1A-4) and these genes were employed to develop the FVC-gene predictor score based on LASSO regression coefficients (Table 4). Hierarchical clustering discriminated FVC decline, while having no association with DLCO decline (Figure 2A). PCA map of the training data confirmed the distinct separation of stable and progressive patients (Figure 2B). The PCA variables factor map aligned the direction of association of individual genes with these groups (Figure 2C).

Table 4

List of 25 Genes, LASSO Regression Coefficient, and Percent of 10-Fold Cross Validation (CV) Support of Each Gene Consists of the FVC-classifier

Gene Symbol*	Gene Description	LASSO Regression Coefficient	% CV Support
<i>APTX</i>	Aprataxin	-1.38	60
<i>ATP6AP1L</i>	ATPase H ⁺ transporting accessory protein 1 like	-2.08	70
<i>CNR2</i>	Cannabinoid receptor 2	-1.066	100
<i>FAM111B</i>	Family with sequence similarity 111 member B	-0.601	50
<i>GABRR1</i>	γ -aminobutyric acid type A receptor rho1 subunit	0.058	50
<i>GPR39</i>	G protein-coupled receptor 39	0.706	50
<i>GYP A</i>	Glycophorin A (MNS blood group)	0.029	60
<i>HBB</i>	Hemoglobin subunit beta	0.037	60
<i>IGLC1</i>	Immunoglobulin lambda constant 1	4.9×10^{-4}	70
<i>ITLN1</i>	Intelectin 1	3.44	100
<i>LINC00319</i>	Long intergenic non-protein coding RNA 319	4.758	100
<i>MAZ</i>	MYC associated zinc finger protein	2.188	80
<i>MSR1</i>	Macrophage scavenger receptor 1	2.157	90

<i>NT5E</i>	5'-nucleotidase ecto	-3.361	80
<i>PAWR</i>	Pro-apoptotic WT1 regulator	-0.515	50
<i>PCDHB15</i>	Protocadherin beta 15	3.521	100
<i>PLA2G4A</i>	Phospholipase A2 group IVA	-0.542	80
<i>PLCL1</i>	Phospholipase C like 1 (inactive)	0.617	50
<i>PNMA5</i>	PNMA family member 5	-0.942	60
<i>RAB3C</i>	RAB3C, RAS oncogene family	-2.218	100
<i>RBM43</i>	RNA binding motif protein 43	0.284	50
<i>RLBP1</i>	Retinaldehyde binding protein 1	0.038	60
<i>SSU72P8</i>	SSU72 pseudogene 8	-1.367	90
<i>TP63</i>	Tumor protein p63	1.71	70
<i>ZNF252P</i>	Zinc finger protein 252, pseudogene	-0.289	70

*FVC-classifier genes were prioritized from COMET training cohort as described in Figures 1A and 1B.

EXAMPLE 3

FVC-gene Predictor Validation

A “cross-platform-gene-match” step to account for transcriptome assay platform differences yielded an overlap of 72% (18/25), 60% (15/25), and 92% (23/25) genes in UChicago, UPMC, and Imperial datasets, respectively (Figure 1B-2).

FVC-gene predictor test performance across validation cohorts is shown in Table 5. Sensitivity and specificity were 1.0 in the training cohort and 0.67-0.8 and 0.78-0.89, respectively in the validation cohorts. Positive predictive values (PPV) ranged from 0.62 to 0.86 and negative predictive values (NPV) ranged from 0.7 to 0.89. ROC analysis revealed AUCs of 0.80, 0.78 and 0.77 in UChicago, UPMC, and Imperial cohort, respectively (Figure 3A). At an anchored specificity of -75%, the sensitivities are 75.0%, 66.7%, and 80.0%, for UChicago, UPMC and Imperial cohorts, respectively. Aggregation of validation cohorts provided a sensitivity/specificity of 74.3%/82.4% with PPV/NPV of 0.74/0.82 (Table 5). AUCs for 5% FVC decline in validation cohorts were 0.74, 0.78, and 0.82 in UChicago, UPMC, and Imperial, respectively (Figures 4A-4C).

EXAMPLE 4

Sensitivity Analysis of Transcriptomic Sampling Timepoints

FVC-gene predictor hierarchical clustering (Figure 5A-5C) and test performance was assessed at other available transcriptomic sampling timepoints in the COMET training

cohort (Table 5). Like the 0-4 month timepoint, patients with FVC decline clustered, while those with DLCO decline did not. ROC analysis of the FVC-gene predictor varied with transcriptome sampling timepoints and demonstrated performance decay with increasing time (Figure 3B). The optimal prediction sensitivity ranged from 0.79 to 0.92, while optimal specificity ranged from 0.67 to 0.85 in the COMET subsets (Table 6). Varying the baseline resulted in worsening test performance as the time from blood draw to FVC measurement decreased (Figure 3C).

Table 5

FVC-gene Predictor Test Performance

ROC Test Results	Training Cohort		Validation Cohorts							
	COMET		UChicago		UPMC		Imperial		Combined	
Observed $\geq$ 10% FV Decline										
	(+)	(-)	(+)	(-)	(+)	(-)	(+)	(-)	(+)	(-)
Predicted (+)	16	0	6	2	8	5	12	2	26	9
Predicted (-)	0	58	2	17	4	18	3	7	9	42
Sensitivity	1		0.75		0.67		0.80		0.74	
Specificity	1		0.89		0.78		0.78		0.82	
PPV	1		0.75		0.62		0.86		0.74	
NPV	1		0.89		0.82		0.70		0.82	
LR (+)	--		6.82		3.05		3.64		4.11	
LR (-)	--		0.28		0.42		0.26		0.32	

ROC= Receiver-Operating-Characteristics; PPV = Positive Predictive Value, NPV = Negative Predictive Value; LR = Likelihood Ratio

Table 6

Receiver-Operating-Characteristic and Area-under-the-Curve Analyses of FVC-gene Predictor for COMET Tand Each COMET Subset Cohort

Cohort (GE sampling duration)	Sample size (Stable /progressive)	*AUC (95% CI)	Optimal Sensitivity/ Specificity	Sensitivity of anchored specificity at -75%
COMET (0-4 mo)	74 (58/16)	1.00	100/100	100
COMET (0-8 mo)	67 (54/13)	0.90 (0.77, 1.02)	92.3/85.2	92.3

COMET (0-12 mo)	59 (45/14)	0.83 (0.7, 0.96)	78.6/77.8	78.6
COMET (4-8 mo)	67 (55/12)	0.78 (0.63, 0.94)	84.6/66.7	69.2
COMET (8-12 mo)	59 (45/14)	0.57 (0.39 0.76)	NA	36.4

*Area-Under-Curve and 95% confidence intervals in ROC analysis of FVC-gene predictor score for $\geq 10\%$ relative decline in FVC percent of predicted.

EXAMPLE 5

Comparative Analysis between FVC-gene Predictor and Cross-sectional Biomarkers

Test performance in predicting FVC decline was compared between the FVC-gene predictor, 4 month change in FVC, and circulating plasma MMP7, POSTN, and CCL18. The FVC-gene predictor outperformed each of these clinical and cross-sectional biomarkers (Figure 6A-6D).

EXAMPLE 6

Pathway and Functional Analyses

Gene Set Enrichment Analysis (GSEA) of 19394 annotated genes in the Δ GE data of COMET training cohort identified several functional pathways. Δ GE of 27 hallmark genes in TGF-beta signaling were higher in progressive patients than in patients with stable disease course (Figure 7A; FDR = 0.036). The gene list, their description, and features related to GSEA score enrichment are shown in Table 7 and included transforming growth factor, beta 1 and its receptor (TGF- β 1 & TGFBR1), and SMAD, mothers against DPP homolog 6 & 7 (SMAD6, SMAD7).

Table 7

List of Hallmark Genes in TGF-beta Signaling Enriched in GSEA Analysis

Gene symbol	Gene description	Rank in gene list	Rank metric score	Running enrichment score*
<i>SERPINE1</i>	serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1	16	1.581	0.064
<i>SKIL</i>	SKI-like	24	1.534	0.127

<i>PMEPA1</i>	prostate transmembrane protein, androgen induced 1	25	1.533	0.190
<i>SMAD7</i>	SMAD, mothers against DPP homolog 7	36	1.406	0.248
<i>SMAD6</i>	SMAD, mothers against DPP homolog 6	84	1.136	0.292
<i>THBS1</i>	thrombospondin 1	117	1.057	0.334
<i>SKI</i>	v-ski sarcoma viral oncogene homolog (avian)	131	1.029	0.376
<i>SMURF1</i>	SMAD specific E3 ubiquitin protein ligase 1	152	0.997	0.416
<i>CTNNB1</i>	catenin (cadherin-associated protein), beta 1, 88kDa	298	0.809	0.442
<i>ENG</i>	endoglin (Osler-Rendu-Weber syndrome 1)	316	0.796	0.474
<i>TGIF1</i>	TGFB induced factor homeobox 1	365	0.764	0.503
<i>TGFBR1</i>	transforming growth factor, beta receptor I (activin A receptor type II-like kinase, 53kDa)	756	0.578	0.506
<i>SMURF2</i>	SMAD specific E3 ubiquitin protein ligase 2	803	0.564	0.527
<i>TRIM33</i>	tripartite motif-containing 33	844	0.551	0.547
<i>BCAR3</i>	breast cancer anti-estrogen resistance 3	969	0.511	0.562
<i>TGFB1</i>	transforming growth factor, beta 1 (Camurati-Engelmann disease)	978	0.509	0.583
<i>ID3</i>	inhibitor of DNA binding 3, dominant negative helix-loop-helix protein	1061	0.490	0.599
<i>BMP2</i>	bone morphogenetic protein 2	1178	0.460	0.611
<i>IFNGR2</i>	interferon gamma receptor 2 (interferon gamma transducer 1)	1322	0.428	0.622

<i>CDK9</i>	cyclin-dependent kinase 9 (CDC2-related kinase)	1377	0.416	0.636
<i>FURIN</i>	furin (paired basic amino acid cleaving enzyme)	1409	0.407	0.651
<i>SMAD3</i>	SMAD, mothers against DPP homolog 3 (Drosophila)	1436	0.400	0.666
<i>SPTBN1</i>	spectrin, beta, non-erythrocytic 1	1453	0.397	0.682
<i>ID1</i>	inhibitor of DNA binding 1, dominant negative helix-loop-helix protein	1642	0.363	0.687
<i>PPM1A</i>	protein phosphatase 1A (formerly 2C), magnesium-dependent, alpha isoform	1662	0.358	0.701
<i>UBE2D3</i>	ubiquitin-conjugating enzyme E2D 3 (UBC4/5 homolog, yeast)	1663	0.358	0.715
<i>NCOR2</i>	nuclear receptor co-repressor 2	1821	0.331	0.721

In contrast, Δ GE values of genes involved in Glycan degradation activity were expressed higher in those in the stable group than progressive group (Figure 7B; Table 8).

Table 8

List of Hallmark Genes in TGF-beta Signaling Enriched in GSEA Analysis

Gene symbol	Gene description	Rank in gene list	Rank metric score	Running enrichment score*
<i>ENGASE</i>	endo-beta-N-acetylglucosaminidase	15208	-0.234	-0.603
<i>NEU1</i>	sialidase 1 (lysosomal sialidase)	16048	-0.315	-0.588
<i>HEXB</i>	hexosaminidase B (beta polypeptide)	16379	-0.348	-0.540
<i>MAN2B2</i>	mannosidase, alpha, class 2B, member 2	17153	-0.448	-0.496
<i>FUCA2</i>	fucosidase, alpha-L-2, plasma	17376	-0.482	-0.417
<i>FUCA1</i>	fucosidase, alpha-L-1, tissue	17657	-0.529	-0.333

<i>AGA</i>	aspartylglucosaminidase	17683	-0.532	-0.234
<i>GBA</i>	glucosidase, beta; acid (includes glucosylceramidase)	18292	-0.681	-0.139
<i>GLBI</i>	galactosidase, beta 1	18945	-0.986	0.012

Functional analysis by TopGene and GOSim of the 25 genes prioritized for the predictor revealed enrichment in “Response to hydrogen peroxide”, “Pulmonary fibrosis” (Table 9), and “receptor-mediated endocytosis”, “positive regulation of fibroblast apoptotic process” (Table 10), respectively.

Table 9

TopGene Functional Enrichment of FVC-Classifer

Function	ID	Name	FDR
GO	GO:0042542	Response to hydrogen peroxide	1.56×10^{-2}
GO	GO:0010035	Response to inorganic substance	1.56×10^{-2}
DisGeNET Curated	C0003504	Aortic Valve Insufficiency	2.31×10^{-2}
DisGeNET Curated	C4025735	Nonspherocytic hemolytic anemia	2.31×10^{-2}
Clinical Variations	cv:C1970028	Susceptibility to malaria	2.31×10^{-2}
OMIM	OMIM:611162	Malaria, susceptibility to	2.31×10^{-2}
DisGeNET Curated	C0034069	Pulmonary Fibrosis	2.31×10^{-2}

Table 10

TopGene Functional Enrichment of FVC-Classifer

Gene Ontology Slim	Adj. p-value	Symbol
Gamma-aminobutyric acid signaling pathway	0.0005	<i>GABRR1; PLCL1</i>
Phagocytosis, engulfment	0.0022	<i>IGLC1; MSRI</i>
AMP catabolic process	0.0026	<i>NT5E</i>
Termination of RNA polymerase II transcription	0.0031	<i>SSU72; MAZ</i>
Nitric oxide transport	0.0039	<i>HBB</i>
Response to nematode	0.0039	<i>ITLN1</i>
Receptor-mediated endocytosis	0.0048	<i>HBB; IGLC1; MSRI</i>

Positive regulation of hydrogen peroxide-mediated programmed cell death	0.0051	<i>PAWR</i>
Platelet activating factor biosynthetic process	0.0051	<i>PLA2G4A</i>
Negative regulation of nitric-oxide synthase activity	0.0090	<i>CNR2</i>
Positive regulation of cholesterol storage	0.0090	<i>MSR1</i>
Positive regulation of fibroblast apoptotic process	0.0090	<i>TP63</i>
Negative regulation of inflammatory response	0.0097	<i>CNR2; NT5E</i>
Response to hydrogen peroxide	0.0097	<i>HBB; PAWR; APTX</i>

EXAMPLE 7

Comparative Analysis between Longitudinal Gene Expression Change and Cross-sectional Gene express in FVC-gene Predictor Modeling

The Coefficient of Variation (CoV) analysis confirmed that longitudinal within-patient Δ GE was more homogenous, and with less within-group variation, than cross-sectional baseline GE data. The majority of the Δ CoV values (difference of CoV between GE and Δ GE) in the MvA (Minus vs Average) plots reside above the CoV1-CoV2 = 0 line in both groups stratified by FVC decline (Figure 8A and 8B; 82.9% and 67.5%, respectively). Using the Δ GE, only requires 16 progressor subjects to achieve a power of 0.9 (1-beta) with an alpha of 0.05 for significance, in contrast to the increase to 63 subjects per group in the cross-sectional baseline GE data (Figure 8C). Only the within-patient Δ GE approach would succeed as in COMET cohort (n = 16 progressors).

Intra-subject CoV was compared between FVC progressor and stable patients in three consecutive transcriptome sampling time points of COMET cohort. Consistently, 60-76% genes demonstrated larger intra-subject CoV in FVC stable than in progressor patients (Figure 8D, grey and black bar, respectively).

Discussion of EXAMPLES

As disclosed herein, a novel, longitudinal gene expression-based predictor of FVC decline was developed. This prediction tool demonstrated good test performance for discriminating progressive and stable patients with IPF across multiple independent cohorts. The performance characteristics support generalizability across varied transcriptome sampling time and durations, and independence of transcriptome assay platform. The

positive and negative predictive values for this tool support its potential use for clinical trial enrichment.

Most IPF clinical trials are designed to detect a clinically meaningful difference in FVC change over time between treatment arms. However, the percentage of patients experiencing FVC decline is highly variable between clinical trial cohorts (Noth et al., 2012; Idiopathic Pulmonary Fibrosis Clinical Research Network et al., 2010; Idiopathic Pulmonary Fibrosis Clinical Research Network et al., 2012; King et al., 2014; Richeldi et al., 2014), requiring relatively large sample sizes to avoid underpowered studies. As such, biomarker-driven enrichment of clinical trial cohorts remains a goal of precision medicine. The ideal biomarker for this purpose should be easily acquired, generalizable across studies and diverse IPF cohorts and reflect underlying pathologic processes. The present data suggest that this tool could potentially serve this role after further refinement in larger cohorts. With a NPV of roughly 80%, this tool could effectively increase the proportion of patients that will experience FVC decline, thereby reducing the number of patients needed for clinical trial enrollment.

While a cross-sectional biomarker would be preferred over one that requires two or more data acquisitions, but as the present data demonstrate, the longitudinal change in a biomarker may better reflect disease activity. The use of short-term Δ GE established a more homogenous transcriptomic predictor of FVC decline than a cross-sectional approach. Additionally, even though the peripheral blood transcriptome was employed for this study, the genes comprising the FVC-gene predictor largely involved known fibrotic pathways, further supporting a reflection of disease activity with this tool. Similar findings have been demonstrated with longitudinal change in circulating plasma biomarkers (Maher et al., 2017).

The ideal timeframe for blood sample acquisition remains unknown. It was shown that this could be done across multiple time points in the COMET cohort and that a year or more delay in sample acquisition still predicted future FVC decline in the UChicago and UPMC registry cohorts, supporting resiliency for variable sampling timepoints. Shorter timeframes were unable to be assessed with these data; however, the shortest possible timeframe for detection of gene expression changes would be ideal for clinical utility. At present, in some embodiments a 4 month sampling duration is a representative duration, but shorter timeframes also fall within the presently disclosed subject matter.

The test performance of FVC-gene predictor disclosed herein against relevant predictors of IPF mortality was explored, including plasma biomarkers and prior FVC decline. MMP7 is a reliable predictor of IPF mortality (Rosas et al., 2008), has been shown to correlate with FVC decline and predicts outcomes in multiple studies (Richards et al., 2012; Hamai et al., 2016) including interstitial lung abnormalities (Armstrong et al., 2017). CCL18, is predictive of outcomes in IPF (Prasse et al., 2009), and has shown prognostic value for absolute change in FVC in two large clinical trial cohorts (Neighbors et al., 2018). COMET investigators previously demonstrated that periostin (POSTN) levels predicted composite progression outcomes (Naik et al., 2012), while others have shown correlation with FVC decline (Tajiri et al., 2015; Ohta et al., 2017). The FVC gene predictor outperformed each of these in detecting future FVC decline.

Results of pathway analyses of all annotated genes associated with FVC decline, and the subset of 25 genes constituting the FVC-predictor confirm pathologic processes involved in IPF, supporting that short-term transcriptomic changes reflect disease activity. The top two pathways identified by GSEA included TGF- β 1 signaling and glycan degradation pathways. Both are directly involved in the fibrotic events of pulmonary fibrosis (Kang et al., 2007) and its pathogenesis (Pardo et al., 2016). Many of the individual genes in the FVC-predictor including TP63, NT5E, FAM111B, HBB, PLA2G4A, MSR1, CNR2, and ITLN1 are also linked to lung fibrosis. TP63 has been reported in the abnormal re-epithelialization and lung remodeling in IPF (Chilosi et al., 2002) while CD73 (NTE5) enhances radiation-induced lung fibrosis in mice, as examples (Wirsdorfer & Jendrosseck, 2016). While the PBMCs are predominantly immune cells, the present findings support that peripheral blood reflect fibrotic signaling in the lungs.

“Loss of transcriptomic robustness” may be explained by the decrease in intra-subject gene expression variation in the FVC progression patients. “Robustness” of a biologic system involves persistence of expression in the face of perturbation (Masel & Siegal, 2009). In essence, alternate pathways other than the perturbed system may be biologically necessary to maintain a healthy response. The CoV analyses disclosed herein showed greater intra-subject gene expression homogeneity in FVC progressor over FVC stable patients. The heterogeneity of Δ GE maintained in patients not experiencing an FVC decline may reflect this preservation of transcriptomic robustness, whereas those with FVC decline may lose this robustness. One cannot infer whether this loss of “robustness” is a cause or a consequence of disease activity, however, this phenomenon establishes a

molecular foundation for application of longitudinal blood transcriptomics in prediction of disease progression.

A strength of the presently disclosed subject matter resides in consistent test performance of the FVC-gene predictor in three independent, international IPF cohorts. While the transcriptome sampling and PFT timepoints were fixed in COMET, such timepoints varied substantially across the IPF registry cohorts (Chicago, UPMC), which more closely approximates clinical practice and illustrates flexibility for clinical application. Equally important, was the diverse transcriptome assay platforms used in these cohorts (RNAseq, Agilent, Affymetrix), which supports generalizability. Another strength of the presently disclosed subject matter was the robustness across FVC decline events. A relative FVC decline of 10% or more is strongly associated with future mortality. However, a relative decline of 5% has also been shown to predict future mortality, so the FVC-gene predictor was also tested for this categorical event and showed similar test performance.

The variability in transcriptome assay platforms led to some data loss across platforms, especially the UPMC platform. Accordingly, absolute FVC-gene predictor score thresholds might not be universally applicable, but rather a modified threshold based on ROCs using the available data has been generated. Gene expression requires normalization prior to downstream data analysis. Batch effects associated with RNA isolation and cDNA library preparation also prohibit uniform scoring and cut-off. Although the predictor has been validated in subsets of the COMET cohort and in external cohorts, the size of each cohort is still relatively small. However, none of the subjects were on FDA approved therapies preventing assessment of responses to drugs.

In conclusion, the FVC-gene predictor included genes with increased or decreased expression over baseline sampling and follow-up suggesting a changing/active disease process. Developing such a blood-derived biomarker for disease activity rather than disease severity could carry implications for clinical trial enrichment and assist clinical decision-making for instituting and maintaining pharmacotherapy. Clinical trial enrichment using such a biomarker could assist with accelerated drug development for this devastating disease. Further refinement of the presently disclosed predictor in larger cohorts on a uniform transcriptome assay platform in conjunction with therapeutic intervention might improve test performance characteristics and facilitate a precision medicine approach in IPF.

Summarily, a training cohort (n = 74) of IPF patients was stratified according to the presence of progressive disease, defined as $\geq 10\%$ relative decline in FVC over 12 months.

Baseline to 4 month within-patient changes in gene expression were correlated with categorical FVC decline. Genes predictive of FVC decline were identified by two-group comparison with false discovery rate <5%, and further prioritized by logistic LASSO regression with $p < 0.05$ and 10-Fold Cross-Validation $\geq 50\%$ support. An FVC-gene score was derived using regression coefficients and assessed using area under the curve (AUC) analysis. The categorical FVC-gene predictor was then applied to three independent validation cohorts with differing transcriptome assay platforms and blood transcriptome sampling times to assess test performance characteristics.

A longitudinally-derived FVC-gene predictor accurately discriminated most patients with stable and progressive IPF across four independent IPF cohorts and demonstrated sensitivity and specificity of 74% and 82% in the combined validation cohort. TGF-beta was the highest-ranking canonical pathway by Gene Set Enrichment Analysis. The use of longitudinal change in gene expression markedly reduced within-group variation compared to a cross-sectional approach. Therefore, a novel predictor of FVC decline developed from longitudinal gene expression accurately discriminated most patients with progressive versus stable IPF. Disease activity may be better reflected in longitudinal over cross-sectional approaches. The resulting FVC-gene predictor may allow enrichment for progressive disease in clinical trials.

REFERENCES

All references listed below, as well as all references cited in the instant disclosure, including but not limited to all patents, patent applications and publications thereof, scientific journal articles, and database entries (e.g., GENBANK® and UniProt biosequence database entries and all annotations available therein) are incorporated herein by reference in their entireties to the extent that they supplement, explain, provide a background for, or teach methodology, techniques, and/or compositions employed herein.

Alexa et al. (2006) Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* 22:1600-1607.

Altschul et al. (1990a) Basic local alignment search tool. *J Mol Biol* 215:403-410.

Altschul et al. (1990b) Protein database searches for multiple alignments. *Proc Natl Acad Sci U S A* 87:14:5509-5513.

Altschul et al. (1997) Gapped BLAST and PSI- BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389-3402.

- American Thoracic Society & European Respiratory Society (2002) American Thoracic Society/European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias. This joint statement of the American Thoracic Society (ATS), and the European Respiratory Society (ERS) was adopted by the ATS board of directors, June 2001 and by the ERS Executive Committee, June 2001. *Am J Respir Crit Care Med* 165:277-304.
- Anders et al. (2014) HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31:166-169.
- Armstrong et al. (2017) Serum Matrix Metalloproteinase-7, Respiratory Symptoms, and Mortality in Community-Dwelling Adults. MESA (Multi-Ethnic Study of Atherosclerosis). *Am J Respir Crit Care Med* 196:1311-1317.
- Ausubel et al. (1995) *Current Protocols in Molecular Biology*, Greene Publishing.
- Ballman et al. (2004) Faster cyclic loess: normalizing RNA arrays via linear models. *Bioinformatics* 20:2778-2786.
- Benjamini & Hochberg (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc* 57:289-300.
- Chilosi et al. (2002) Abnormal re-epithelialization and lung remodeling in idiopathic pulmonary fibrosis: the role of deltaN-p63. *Lab Invest* 82:1335-1345.
- Dobin et al. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15-21.
- Devereux et al. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucl Acids Res* 12:387-395.
- du Bois et al. (2011) Forced vital capacity in patients with idiopathic pulmonary fibrosis: test properties and minimal clinically important difference. *Am J Respir Crit Care Med* 184:1382-1389.
- Flaherty et al. (2004) Idiopathic interstitial pneumonia: what is the effect of a multidisciplinary approach to diagnosis? *Am J Respir Crit Care Med* 170:904-910.
- Flaherty et al. (2007) Idiopathic interstitial pneumonia: do community and academic physicians agree on diagnosis? *Am J Respir Crit Care Med* 175:1054-1060.
- Friedman et al. (2010) Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* 33:1-22.
- Gait (1984) *Oligonucleotide Synthesis: A Practical Approach*, IRL Press, Oxford, England.
- Gautier et al. (2004) affy--analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 20:307-315.

- Glover (1985) *DNA Cloning: a Practical Approach*. Oxford Press, Oxford.
- Greene et al. (2002) Serum surfactant proteins-A and -D as biomarkers in idiopathic pulmonary fibrosis. *Eur Respir J* 19:439-446.
- Gross & Mienhofer (eds.) (1981) The Peptides, Vol. 3. Academic Press, New York, New York, United States of America, pp. 3-88.
- Gu et al. (2016) Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32:2847-2849.
- Harlow & Lane (1988) Antibodies, a Laboratory Manual, Cold Spring Harbor Laboratory Publications, Cold Spring Harbor, New York, United States of America.
- Hamai et al. (2016) Comparative Study of Circulating MMP-7, CCL18, KL-6, SP-A, and SP-D as Disease Markers of Idiopathic Pulmonary Fibrosis. *Dis Markers* 2016; 2016: 4759040.
- Herazo-Maya et al. (2013) Peripheral blood mononuclear cell gene expression profiles predict poor outcome in idiopathic pulmonary fibrosis. *Sci Transl Med* 5:205ra136.
- Herazo-Maya et al. (2017) Validation of a 52-gene risk profile for outcome prediction in patients with idiopathic pulmonary fibrosis: an international, multicentre, cohort study. *Lancet Respir Med* 5:857-868.
- Hoffmann-Vold et al. (2016) High Level of Chemokine CCL18 Is Associated With Pulmonary Function Deterioration, Lung Fibrosis Progression, and Reduced Survival in Systemic Sclerosis. *Chest* 150:299-306.
- Huang et al. (2017) Microbes Are Associated with Host Innate Immune Response in Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med* 196:208-219.
- Husson et al. (2010) Exploratory Multivariate Analysis by Example Using R. CRC Press, Taylor & Francis Group, Boca Raton, Florida, United States of America.
- Hyslop & White (2009) Estimating precision using duplicate measurements. *J Air Waste Manag Assoc* 59:1032-1039.
- Idiopathic Pulmonary Fibrosis Clinical Research Network et al. (2010) A controlled trial of sildenafil in advanced idiopathic pulmonary fibrosis. *N Engl J Med* 363:620-628.
- Idiopathic Pulmonary Fibrosis Clinical Research Network et al. (2012) Prednisone, azathioprine, and N-acetylcysteine for pulmonary fibrosis. *N Engl J Med* 366:1968-1977.
- Irizarry et al. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4:249-264.

- Jegal et al. (2005) Physiology is a stronger predictor of survival than pathology in fibrotic interstitial pneumonia. *Am J Respir Crit Care Med* 171:639-644.
- Kaimal et al. (2010) ToppCluster: a multiple gene list feature analyzer for comparative enrichment clustering and network-based dissection of biological systems. *Nucleic Acids Res* 38:W96-102.
- Kaner et al. (2019) Design of IPF Clinical Trials in the Era of Approved Therapies. *Am J Respir Crit Care Med* 200:133-139.
- Kang et al. (2007) Transforming growth factor (TGF)-beta1 stimulates pulmonary fibrosis and inflammation via a Bax-dependent, bid-activated pathway that involves matrix metalloproteinase-12. *J Biol Chem* 282:7723-7732.
- Karimi-Shah & Chowdhury (2015) Forced vital capacity in idiopathic pulmonary fibrosis--FDA review of pirfenidone and nintedanib. *N Engl J Med* 372:1189-1191.
- Karlin & Altschul (1990) Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc Natl Acad Sci U S A* 87:2264-2268.
- Karlin & Altschul (1993) Applications and statistics for multiple high-scoring segments in molecular sequences. *Proc Natl Acad Sci U S A* 90:5873-5877.
- King et al. (2014) A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *N Engl J Med* 370:2083-2092.
- Law et al. (2014) voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 15:R29.
- Ley et al. (2014) Molecular biomarkers in idiopathic pulmonary fibrosis. *Am J Physiol Lung Cell Mol Physiol* 307:L681-691.
- Ley et al. (2016) Predictors of Mortality Poorly Predict Common Measures of Disease Progression in Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med* 194:711-718.
- Li (2008) Automating dChip: toward reproducible sharing of microarray data analysis. *BMC Bioinformatics* 9:231.
- López-Ratón et al. (2014) OptimalCutpoints: An R Package for Selecting Optimal Cutpoints in Diagnostic Tests. *Journal of Statistical Software* 61:1-36.
- Love et al. (2014) Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol* 15:550.

- Macintyre et al. (2005) Standardisation of the single-breath determination of carbon monoxide uptake in the lung. *Eur Respir J* 26:720-735.
- Maher et al. (2017) An epithelial biomarker signature for idiopathic pulmonary fibrosis: an analysis from the multicentre PROFILE cohort study. *Lancet Respir Med* 5:946-955.
- Masel & Siegal (2009) Robustness: mechanisms and consequences. *Trends Genet* 25:395-403.
- Meyer (2014) Support vector machines: the interface to libsvm in package e1071.
- Miller et al. (2005a) General considerations for lung function testing. *Eur Respir J* 26:153-161.
- Miller et al. (2005b) Standardisation of spirometry. *Eur Respir J* 26:319-338.
- Mortazavi et al. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5:621-628.
- Naik et al. (2012) Periostin promotes fibrosis and predicts progression in patients with idiopathic pulmonary fibrosis. *Am J Physiol Lung Cell Mol Physiol* 303:L1046-1056.
- Neighbors et al. (2018) Prognostic and predictive biomarkers for patients with idiopathic pulmonary fibrosis treated with pirfenidone: post-hoc assessment of the CAPACITY and ASCEND trials. *Lancet Respir Med* 6:615-626.
- Noth et al. (2012) Idiopathic Pulmonary Fibrosis Clinical Research N. A placebo-controlled randomized trial of warfarin in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 186:88-95.
- O'Dwyer et al. (2017) The peripheral blood proteome signature of idiopathic pulmonary fibrosis is distinct from normal and is associated with novel immunological processes. *Sci Rep* 7:46560.
- Ohta et al. (2017) The usefulness of monomeric periostin as a biomarker for idiopathic pulmonary fibrosis. *PLoS One* 12:e0174547.
- Pardo et al. (2016) Role of matrix metalloproteinases in the pathogenesis of idiopathic pulmonary fibrosis. *Respir Res* 17:23.
- Peljto et al. (2013) Association between the MUC5B promoter polymorphism and survival in patients with idiopathic pulmonary fibrosis. *JAMA* 309:2232-2239.
- Prasse et al. (2009) Serum CC-chemokine ligand 18 concentration predicts outcome in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 179: 717-723.

- Raghu et al. (2011) An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med* 183:788-824.
- Richards et al. (2012) Plasma proteins for risk prediction in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 185:1329-1330.
- Richeldi et al. (2014) Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med* 370:2071-2082.
- Robin et al. (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12:77.
- Robinson & Oshlack (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* 11:R25.
- Robinson et al. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139-140.
- Roe et al. (1996) DNA Isolation and Sequencing: Essential Techniques, John Wiley, New York, New York, United States of America.
- Rosas et al. (2008) MMP1 and MMP7 as potential peripheral blood biomarkers in idiopathic pulmonary fibrosis. *PLoS Med* 5:e93.
- Sambrook et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Publications, Cold Spring Harbor, New York, United States of America.
- Schmidt et al. (2014) Predicting pulmonary fibrosis disease course from past trends in pulmonary function. *Chest* 145:579-585.
- Simon et al. (2011) Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent. *J Stat Softw* 39:1-13.
- Smyth (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 3:Article3.
- Subramanian et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102:15545-15550.
- Suykens & Vandewalle (1999) Least Squares Support Vector Machine Classifiers. *Neural Processing Letters* 9:293-300.
- Tajiri et al. (2015) Serum level of periostin can predict long-term outcome of idiopathic pulmonary fibrosis. *Respir Investig* 53:73-81.

- Tibshirani et al. (2012) Strong rules for discarding predictors in lasso-type problems. J R Stat Soc Series B Stat Methodol 74:245-266.
- Trapnell et al. (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25:1105-1111.
- U.S. Patent Application Publication Nos. 2010/0120097; 2011/0189679; 2014/0113333; 2015/0307874; 2018/0064695; 2018/0169084; 2019/0030012; 2019/0282565.
- U.S. Patent Nos. 3,974,281; 5,800,992; 6,004,755; 6,013,449; 6,020,135; 6,033,860; 6,040,138; 6,177,248; 6,251,601; 6,309,822; 6,762,180; 7,824,856; 8,592,462; 9,884,802; 9,920,367; 10,028,966; 10,105,365; 10,227,584; each of which is incorporated by reference in its entirety.
- Wirsdorfer & Jendrossek (2016) The Role of Lymphocytes in Radiotherapy-Induced Adverse Late Effects in the Lung. Front Immunol 7:591.

While the presently disclosed subject matter has been disclosed with reference to specific embodiments, it is apparent that other embodiments and variations of the presently disclosed subject matter may be devised by others skilled in the art without departing from the true spirit and scope of the presently disclosed subject matter.

CLAIMS

What is claimed is:

1. A method for generating a prognostic signature for a subject diagnosed with Idiopathic Pulmonary Fibrosis (IPF) with respect to decline in lung Forced Vital Capacity (FVC), the method comprising:
 - (a) determining a first expression level for one or more genes selected from the group consisting of *ALDH4A1*, *APTX*, *ATP6AP1L*, *CCNB1*, *CNR2*, *DNAJC17*, *DTWD1*, *FAM111B*, *GABRR1*, *GPR39*, *GYPA*, *HBB*, *HLA-DPB1*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MRPL35*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PLCL1*, *PNMA5*, *RAB3C*, *RBM43*, *RLBP1*, *SESN3*, *SLC25A37*, *SSU72P8*, *TP63*, *WDR17*, *ZNF252P*, and *ZNF582* in a first biological sample obtained from the subject diagnosed with IPF to establish a baseline expression level for the one or more genes;
 - (b) determining a second expression level for the one or more genes in a second biological sample obtained from the subject, wherein the first and second biological samples comprise peripheral blood mononuclear cells (PBMCs) and/or nucleic acids extracted from PBMCs; and
 - (c) comparing the first and second expression levels for the one or more genes; wherein the comparing provides a prognostic signature for the subject with respect to decline in lung FVC within two years from the time that the first biological sample was obtained from the subject.
2. The method of claim 1, comprising determining first and second expression levels for a set of genes selected from the group consisting of:
 - (a) *APTX*, *CNR2*, *GYPA*, *ITLN1*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PLA2G4A*, and *PNMA5*;
 - (b) *APTX*, *ATP6AP1L*, *ITLN1*, *LINC00319*, *MAZ*, *MSR1*, *NT5E*, *PCDHB15*, *RAB3C*, *SSU72P8*, and *TP63*;
 - (c) *APTX*, *CNR2*, *GABRR1*, *GPR39*, *GYPA*, *HBB*, *ITLN1*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PNMA5*, *RLBP1*, and *SSU72P8*; and
 - (d) *APTX*, *ATP6AP1L*, *CNR2*, *FAM111B*, *GABRR1*, *GPR39*, *GYPA*, *HBB*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PLCL1*, *PNMA5*, *RAB3C*, *RBM43*, *RLBP1*, *SSU72P8*, *TP63*, and *ZNF252P*.

3. The method of claim 1 or claim 2, comprising determining first and second expression levels for each of *APTX*, *ATP6AP1L*, *CNR2*, *FAM111B*, *GABRR1*, *GPR39*, *GYPA*, *HBB*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PLCL1*, *PNMA5*, *RAB3C*, *RBM43*, *RLBP1*, *SSU72P8*, *TP63*, and *ZNF252P*.
4. The method of any one of the preceding claims, wherein the second biological sample is obtained from the subject at a time from about 4 to about 12 months subsequent to when the first biological sample was obtained from the subject.
5. The method of any one of the preceding claims, wherein the subject is a human.
6. The method of any one of the preceding claims, wherein one or both determining step comprises a technique selected from the group consisting of RNA-seq analysis, quantitative polymerase chain reaction (PCR) including quantitative reverse transcription PCR (qRT-PCR), and the use of a nucleic acid or protein array, or any combination thereof.
7. The method of any one of the preceding claims, wherein the comparing comprises comparing a normalized expression level for each gene in the first biological sample to a normalized expression level for each gene in the second biological sample to generate a fold-increase and/or a fold-decrease in the second biological sample relative to the first biological sample for each gene.
8. The method of claim 7, wherein the comparing comprises summing each fold-increase and/or fold-decrease to produce an FVC-gene predictor score for the subject.
9. The method of claim 8, wherein the summing is performed after multiplying each fold-increase and/or fold-decrease by a weighting value to produce a weighted FVC-gene predictor score for the subject.
10. A method for classifying a subject diagnosed with Idiopathic Pulmonary Fibrosis (IPF) as being at risk for a decline in lung Forced Vital Capacity (FVC), the method comprising:
 - (a) determining a first expression level for one or more genes selected from the group consisting of *ALDH4A1*, *APTX*, *ATP6AP1L*, *CCNB1*, *CNR2*, *DNAJC17*, *DTWD1*, *FAM111B*, *GABRR1*, *GPR39*, *GYPA*, *HBB*, *HLA-DPB1*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MRPL35*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PLCL1*, *PNMA5*, *RAB3C*, *RBM43*, *RLBP1*, *SESN3*,

SLC25A37, *SSU72P8*, *TP63*, *WDR17*, *ZNF252P*, and *ZNF582* in a first biological sample obtained from the subject diagnosed with IPF to establish a baseline expression level for the one or more genes;

- (b) determining a second expression level for the one or more genes in a second biological sample obtained from the subject, wherein the first and second biological samples comprise peripheral blood mononuclear cells (PBMCs) and/or nucleic acids extracted from PBMCs; and
- (c) comparing the first and second expression levels for the one or more genes to create an FVC-gene predictor score;

wherein if the FVC-gene predictor score is greater than or equal to a pre-selected value, the patient is classified as being at risk for a decline in lung FVC within two years from the time that the first biological sample was obtained from the subject.

11. The method of claim 10, wherein the comparing comprises comparing a normalized expression level for each gene in the first biological sample to a normalized expression level for each gene in the second biological sample to generate a fold-increase and/or a fold-decrease in the second biological sample relative to the first biological sample for each gene.
12. The method of claim 11, wherein the comparing further comprises summing each fold-increase and/or fold-decrease to produce an FVC-gene predictor score for the subject.
13. The method of claim 12, wherein the summing is performed after multiplying each fold-increase and/or fold-decrease by a weighting value to produce a weighted FVC-gene predictor score for the subject.
14. The method of any one of claims 10-13, comprising determining first and second expression levels for a set of genes selected from the group consisting of:
 - (a) *APTX*, *CNR2*, *GYP A*, *ITLN1*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PLA2G4A*, and *PNMA5*;
 - (b) *APTX*, *ATP6AP1L*, *ITLN1*, *LINC00319*, *MAZ*, *MSR1*, *NT5E*, *PCDHB15*, *RAB3C*, *SSU72P8*, and *TP62*;
 - (c) *APTX*, *CNR2*, *GABRR1*, *GPR39*, *GYP A*, *HBB*, *ITLN1*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PNMA5*, *RLBP1*, and *SSU72P8*; and
 - (d) *APTX*, *ATP6AP1L*, *CNR2*, *FAM111B*, *GABRR1*, *GPR39*, *GYP A*, *HBB*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*,

PLA2G4A, PLCL1, PNMA5, RAB3C, RBM43, RLBP1, SSU72P8, TP63, and ZNF252P.

15. The method of any one of claims 10-14, comprising determining first and second expression levels for each of *APTX, ATP6AP1L, CNR2, FAM111B, GABRR1, GPR39, GYPA, HBB, IGLC1, ITLN1, LINC00319, MAZ, MSR1, NT5E, PAWR, PCDHB15, PLA2G4A, PLCL1, PNMA5, RAB3C, RBM43, RLBP1, SSU72P8, TP63, and ZNF252P*.
16. The method of any one of claims 10-15, wherein the second biological sample is obtained from the subject at a time from about 4 to about 12 months subsequent to when the first biological sample was obtained from the subject.
17. The method of any one of claims 10-16, wherein the subject is a human.
18. The method of any one of claims 10-17, wherein one or both determining steps comprises a technique selected from the group consisting of RNA-seq analysis, quantitative polymerase chain reaction (PCR) including quantitative reverse transcription PCR (qRT-PCR), and the use of a nucleic acid or protein array, or any combination thereof.
19. A method for identifying and treating a subject diagnosed with Idiopathic Pulmonary Fibrosis (IPF) at risk for a decline in lung Forced Vital Capacity (FVC), the method comprising:
 - (a) determining a first expression level for one or more genes selected from the group consisting of *ALDH4A1, APTX, ATP6AP1L, CCNB1, CNR2, DNAJC17, DTWD1, FAM111B, GABRR1, GPR39, GYPA, HBB, HLA-DPB1, IGLC1, ITLN1, LINC00319, MAZ, MRPL35, MSR1, NT5E, PAWR, PCDHB15, PLA2G4A, PLCL1, PNMA5, RAB3C, RBM43, RLBP1, SESN3, SLC25A37, SSU72P8, TP63, WDR17, ZNF252P, and ZNF582* in a first biological sample obtained from the subject diagnosed with IPF to establish a baseline expression level for the one or more genes;
 - (b) determining a second expression level for the one or more genes in a second biological sample obtained from the subject, wherein the first and second biological samples comprise peripheral blood mononuclear cells (PBMCs) and/or nucleic acids extracted from PBMCs;
 - (c) comparing the first and second expression levels for the one or more genes to create an FVC-gene predictor score; and

- (d) if the FVC-gene predictor score is greater than or equal to a pre-selected value, treating the subject with a treatment selected from the group consisting of lung transplantation and a drug therapy.
- 20. The method of claim 19, wherein the drug therapy comprises administering to the subject a pharmaceutical composition comprising pirfenidone, nintedanib, or a combination thereof in an amount and via a route of administration effective to delay or prevent the development of FVC decline in the subject.
- 21. The method of claim 19 or claim 20, wherein the comparing comprises comparing a normalized expression level for each gene in the first biological sample to a normalized expression level for each gene in the second biological sample to generate a fold-increase and/or a fold-decrease in the second biological sample relative to the first biological sample for each gene.
- 22. The method of claim 21, wherein the comparing further comprises summing each fold-increase and/or fold-decrease to produce an FVC-gene predictor score for the subject.
- 23. The method of claim 22, wherein the summing is performed after multiplying each fold-increase and/or fold-decrease by a weighting value to produce a weighted FVC-gene predictor score for the subject.
- 24. The method of any one of claims 19-23, comprising determining first and second expression levels for a set of genes selected from the group consisting of:
 - (a) *APTX*, *CNR2*, *GYP A*, *ITLN1*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PLA2G4A*, and *PNMA5*;
 - (b) *APTX*, *ATP6AP1L*, *ITLN1*, *LINC00319*, *MAZ*, *MSR1*, *NT5E*, *PCDHB15*, *RAB3C*, *SSU72P8*, and *TP62*;
 - (c) *APTX*, *CNR2*, *GABRR1*, *GPR39*, *GYP A*, *HBB*, *ITLN1*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PNMA5*, *RLBP1*, and *SSU72P8*; and
 - (d) *APTX*, *ATP6AP1L*, *CNR2*, *FAM111B*, *GABRR1*, *GPR39*, *GYP A*, *HBB*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PLCL1*, *PNMA5*, *RAB3C*, *RBM43*, *RLBP1*, *SSU72P8*, *TP63*, and *ZNF252P*.
- 25. The method of any one of claims 19-24, comprising determining first and second expression levels for each of *APTX*, *ATP6AP1L*, *CNR2*, *FAM111B*, *GABRR1*, *GPR39*, *GYP A*, *HBB*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MSR1*, *NT5E*, *PAWR*,

- PCDHB15, PLA2G4A, PLCL1, PNMA5, RAB3C, RBM43, RLBP1, SSU72P8, TP63, and ZNF252P.*
26. The method of any one of claims 19-25, wherein the second biological sample is obtained from the subject at a time from about 4 to about 12 months subsequent to when the first biological sample was obtained from the subject.
 27. The method of any one of claims 19-26, wherein the subject is a human.
 28. The method of any one of claims 19-27, wherein one or both determining steps comprises a technique selected from the group consisting of RNA-seq analysis, quantitative polymerase chain reaction (PCR) including quantitative reverse transcription PCR (qRT-PCR), and the use of a nucleic acid or protein array, or any combination thereof.
 29. A method for monitoring the progress of a treatment in an Idiopathic Pulmonary Fibrosis (IPF) patient whose is experiencing a decline in lung Forced Vital Capacity (FVC), the method comprising:
 - (a) determining a first expression level for one or more genes selected from the group consisting of *ALDH4A1, APTX, ATP6AP1L, CCNB1, CNR2, DNAJC17, DTWD1, FAM111B, GABRR1, GPR39, GYPA, HBB, HLA-DPB1, IGLC1, ITLN1, LINC00319, MAZ, MRPL35, MSR1, NT5E, PAWR, PCDHB15, PLA2G4A, PLCL1, PNMA5, RAB3C, RBM43, RLBP1, SESN3, SLC25A37, SSU72P8, TP63, WDR17, ZNF252P, and ZNF582* in a first biological sample obtained from the patient to establish a baseline expression level for the one or more genes;
 - (b) determining a second expression level for the one or more genes in a second biological sample obtained from the patient at a subsequent time point, wherein the first and second biological samples comprise peripheral blood mononuclear cells (PBMCs) and/or nucleic acids extracted from PBMCs; and
 - (c) comparing the first and second expression levels for the one or more genes, wherein the comparing step is indicative of the progress of the treatment in the patient.
 30. The method of claim 29, wherein the treatment comprises administering to the patient a pharmaceutical composition comprising pirfenidone, nintedanib, or a combination.

31. The method of claim 29 or claim 30, wherein the comparing comprises comparing a normalized expression level for each gene in the first biological sample to a normalized expression level for each gene in the second biological sample to generate a fold-increase and/or a fold-decrease in the second biological sample relative to the first biological sample for each gene.
32. The method of claim 31, wherein the comparing further comprises summing each fold-increase and/or fold-decrease to produce an FVC-gene predictor score for the patient.
33. The method of claim 32, wherein the summing is performed after multiplying each fold-increase and/or fold-decrease by a weighting value to produce a weighted FVC-gene predictor score for the patient.
34. The method of any one of claims 29-33, comprising determining first and second expression levels for a set of genes selected from the group consisting of:
 - (a) *APTX*, *CNR2*, *GYP A*, *ITLN1*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PLA2G4A*, and *PNMA5*;
 - (b) *APTX*, *ATP6AP1L*, *ITLN1*, *LINC00319*, *MAZ*, *MSR1*, *NT5E*, *PCDHB15*, *RAB3C*, *SSU72P8*, and *TP62*;
 - (c) *APTX*, *CNR2*, *GABRR1*, *GPR39*, *GYP A*, *HBB*, *ITLN1*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PNMA5*, *RLBP1*, and *SSU72P8*; and
 - (d) *APTX*, *ATP6AP1L*, *CNR2*, *FAM111B*, *GABRR1*, *GPR39*, *GYP A*, *HBB*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PLCL1*, *PNMA5*, *RAB3C*, *RBM43*, *RLBP1*, *SSU72P8*, *TP63*, and *ZNF252P*.
35. The method of any one of claims 29-34, comprising determining first and second expression levels for each of *APTX*, *ATP6AP1L*, *CNR2*, *FAM111B*, *GABRR1*, *GPR39*, *GYP A*, *HBB*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PLCL1*, *PNMA5*, *RAB3C*, *RBM43*, *RLBP1*, *SSU72P8*, *TP63*, and *ZNF252P*.
36. The method of any one of claims 29-35, wherein the second biological sample is obtained from the patient at a time subsequent to when the first biological sample was obtained from the patient selected from the group consisting of about 1 week, about 2 weeks, about 4 weeks, about 6 weeks, about 2 months, about 3 months, about 4 months, about 5 months, about 6 months, or longer than six months.

37. The method of any one of claims 29-36, wherein the patient is a human.
38. The method of any one of claims 29-37, wherein one or both determining steps comprises a technique selected from the group consisting of RNA-seq analysis, quantitative polymerase chain reaction (PCR) including quantitative reverse transcription PCR (qRT-PCR), and the use of a nucleic acid or protein array, or any combination thereof.
39. The method of any one of claims 29-38, further comprising:
- (d) determining a one or more subsequent expression levels for one or more genes selected from the group consisting of *ALDH4A1*, *APTX*, *ATP6AP1L*, *CCNB1*, *CNR2*, *DNAJC17*, *DTWD1*, *FAM111B*, *GABRR1*, *GPR39*, *GYPA*, *HBB*, *HLA-DPB1*, *IGLC1*, *ITLN1*, *LINC00319*, *MAZ*, *MRPL35*, *MSR1*, *NT5E*, *PAWR*, *PCDHB15*, *PLA2G4A*, *PLCL1*, *PNMA5*, *RAB3C*, *RBM43*, *RLBP1*, *SESN3*, *SLC25A37*, *SSU72P8*, *TP63*, *WDR17*, *ZNF252P*, and *ZNF582* in one or more subsequently isolated biological samples obtained from the patient; and
 - (e) comparing the first, second, and one or more subsequent expression levels for the one or more genes,
- wherein the comparing step is indicative of the progress of the treatment in the patient.

1/15

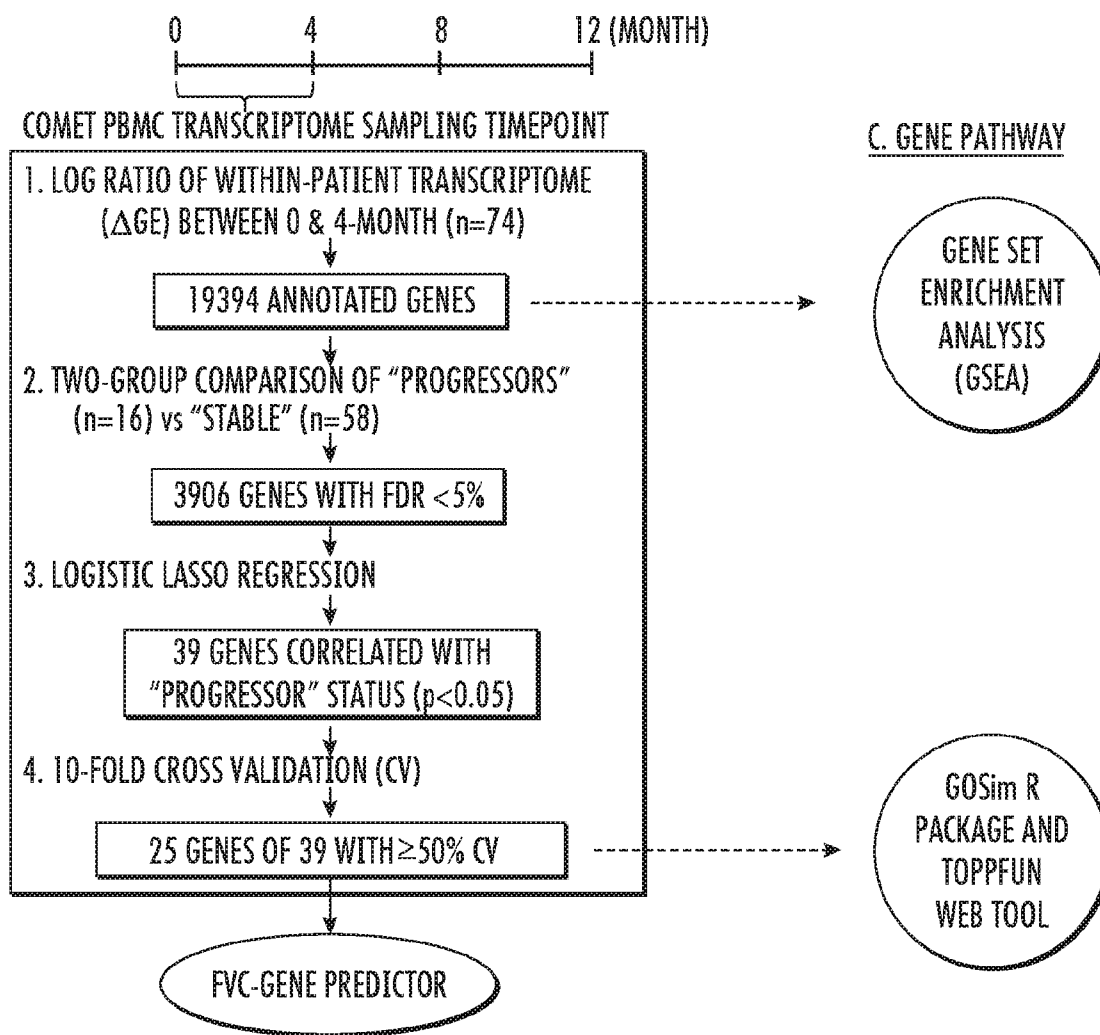


FIG. 1A

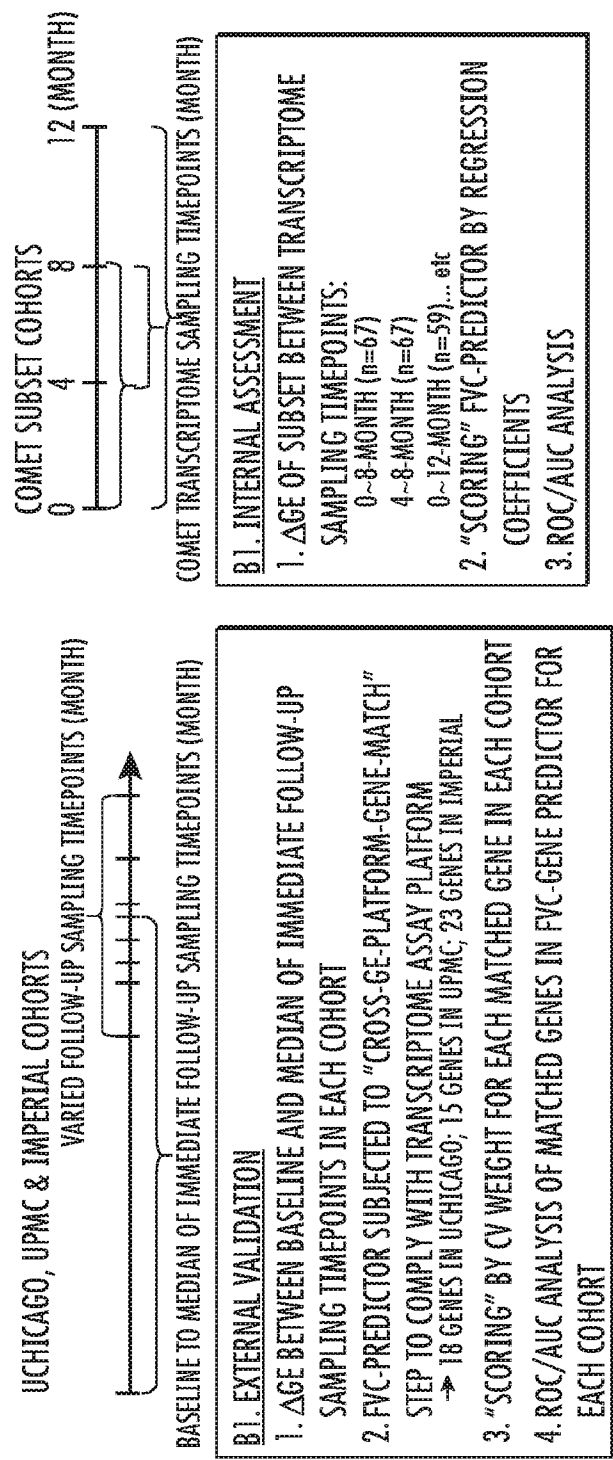
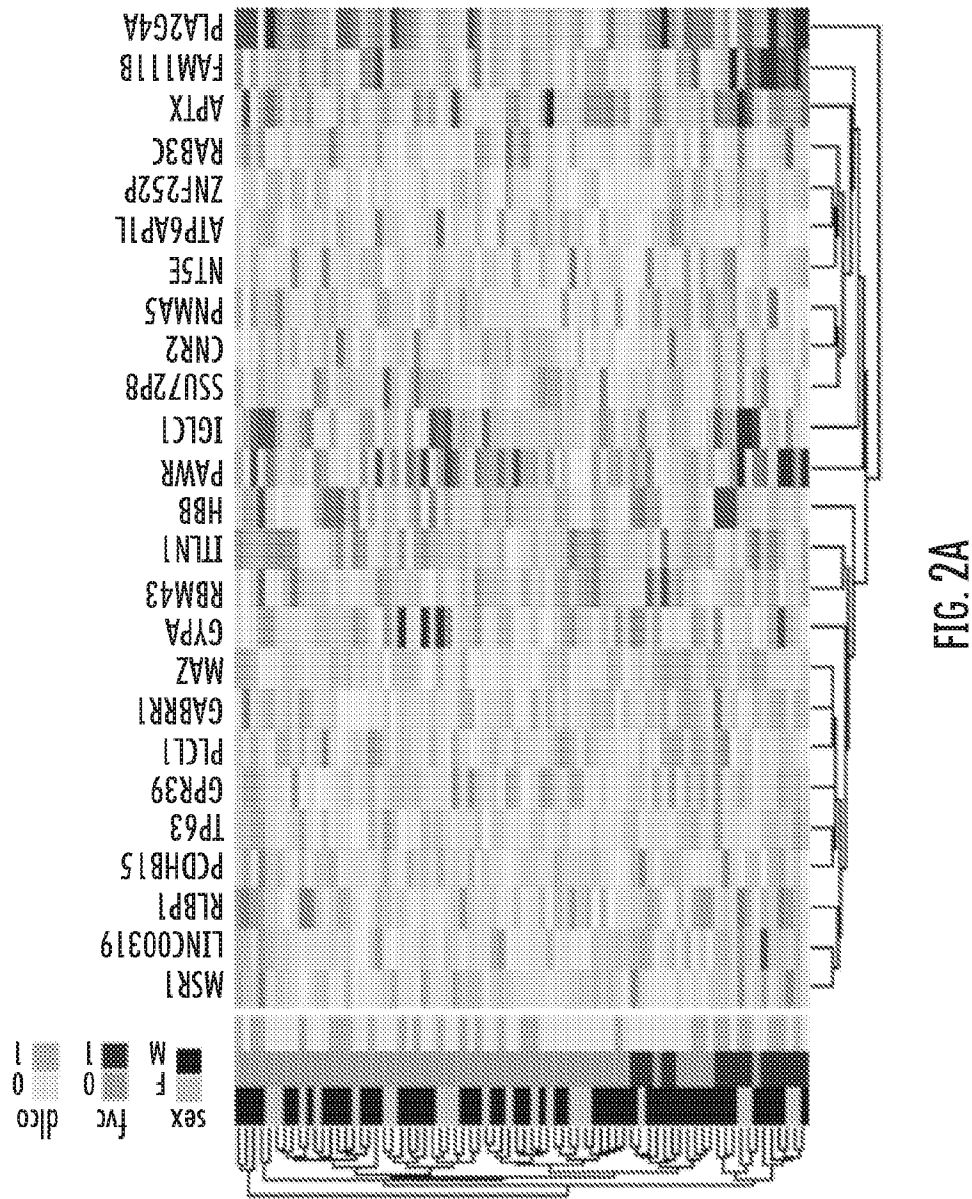


FIG. 1B



4/15

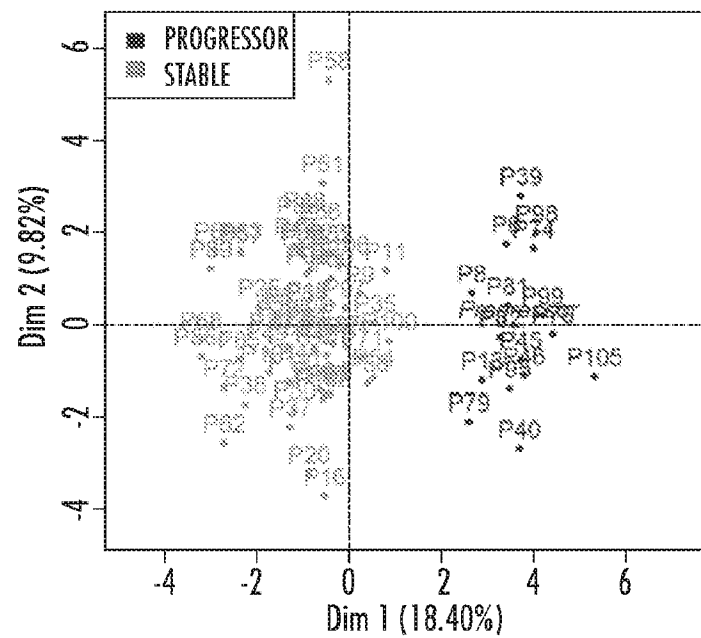


FIG. 2B

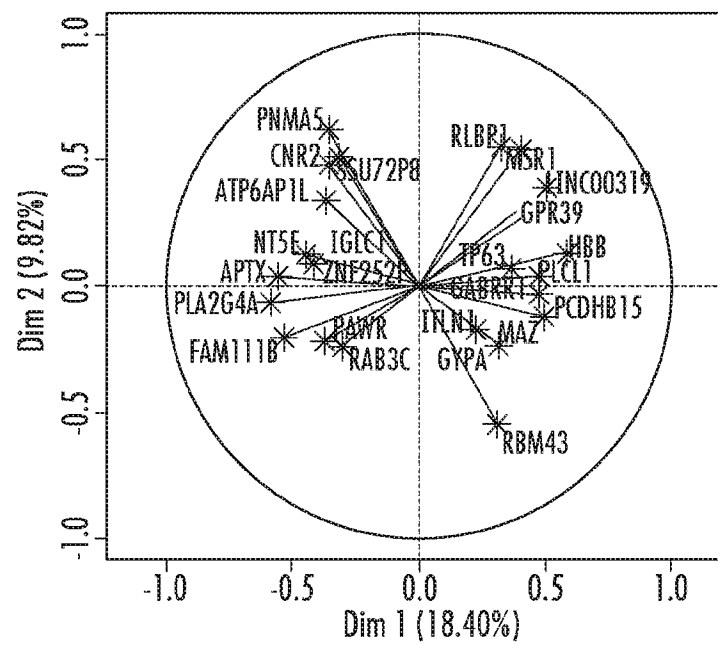


FIG. 2C

5/15

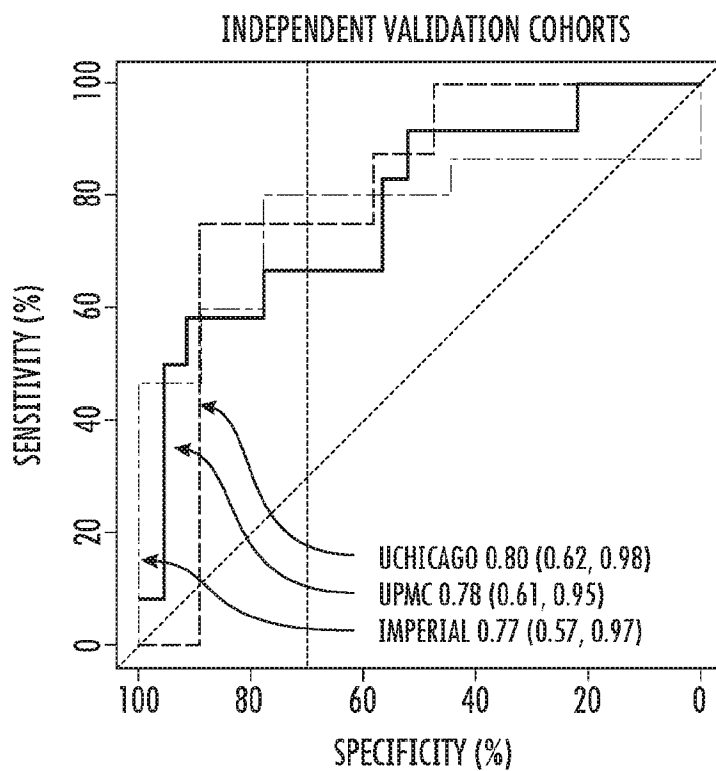


FIG. 3A

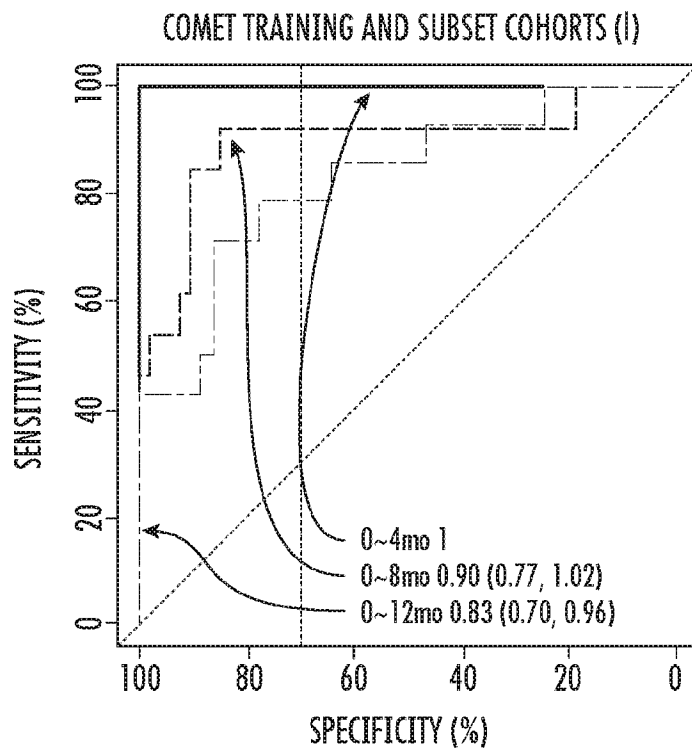


FIG. 3B

6/15

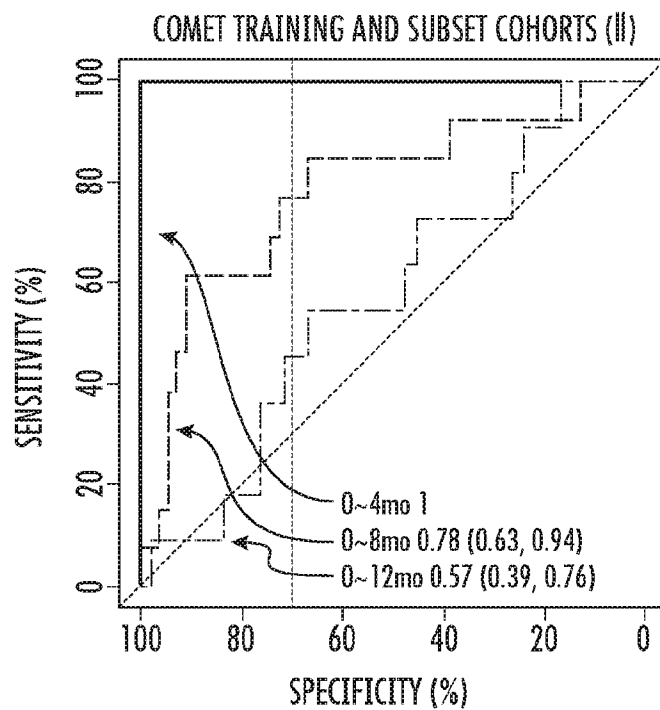


FIG. 3C

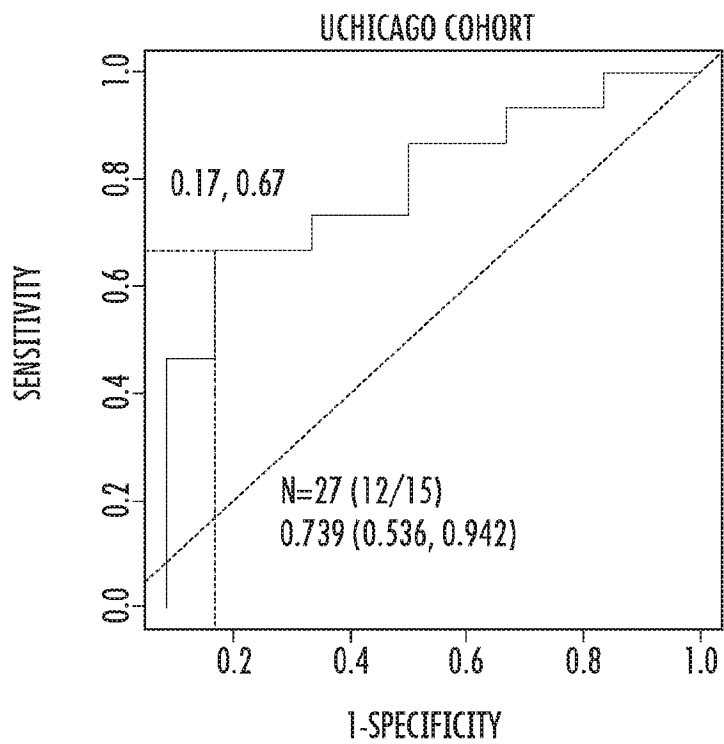


FIG. 4A

7/15

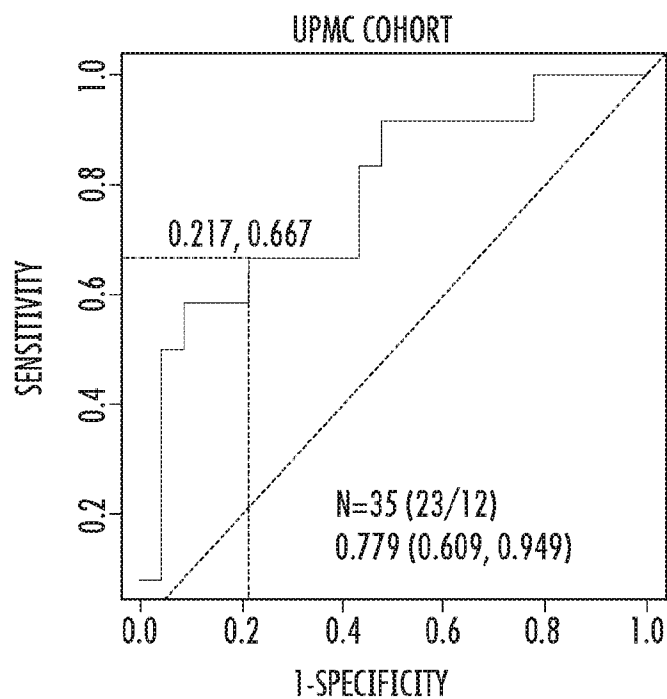


FIG. 4B

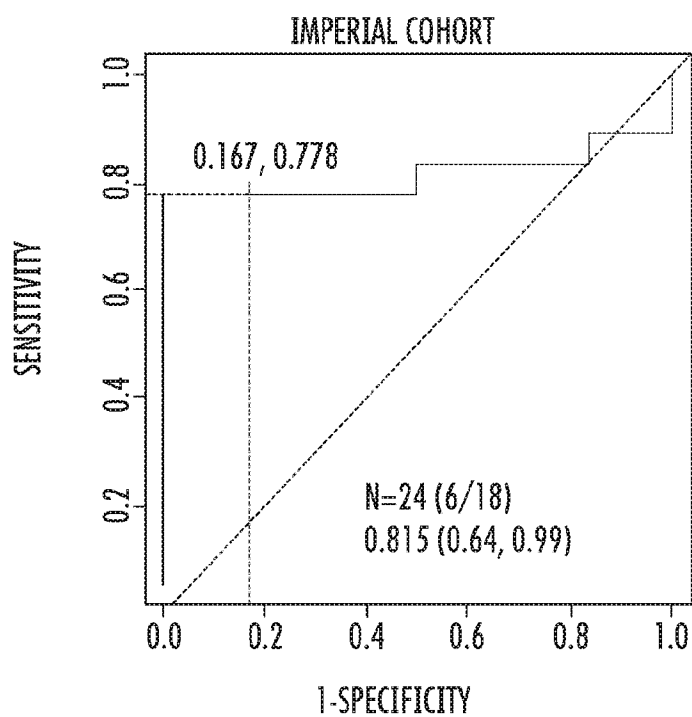


FIG. 4C

8/15

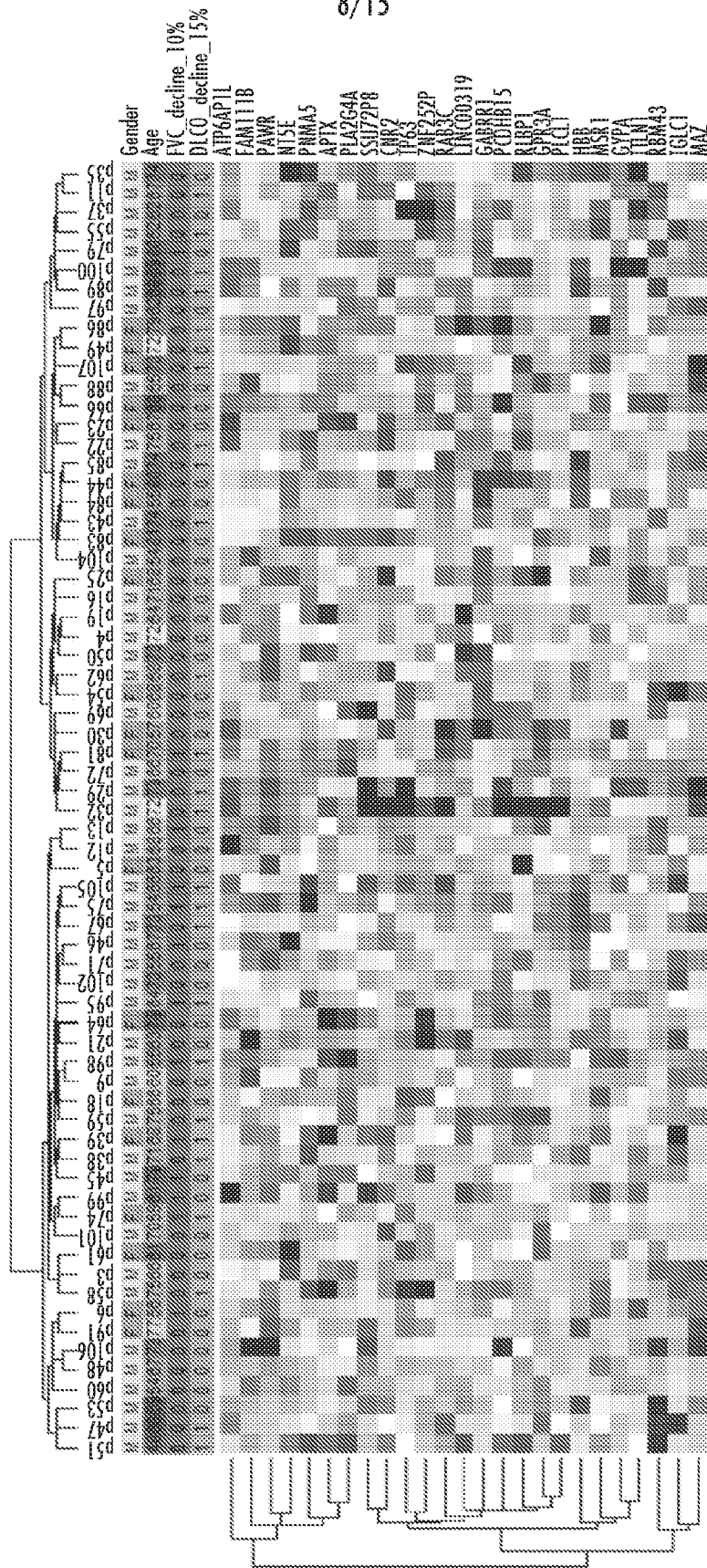


FIG. 5A

9/15

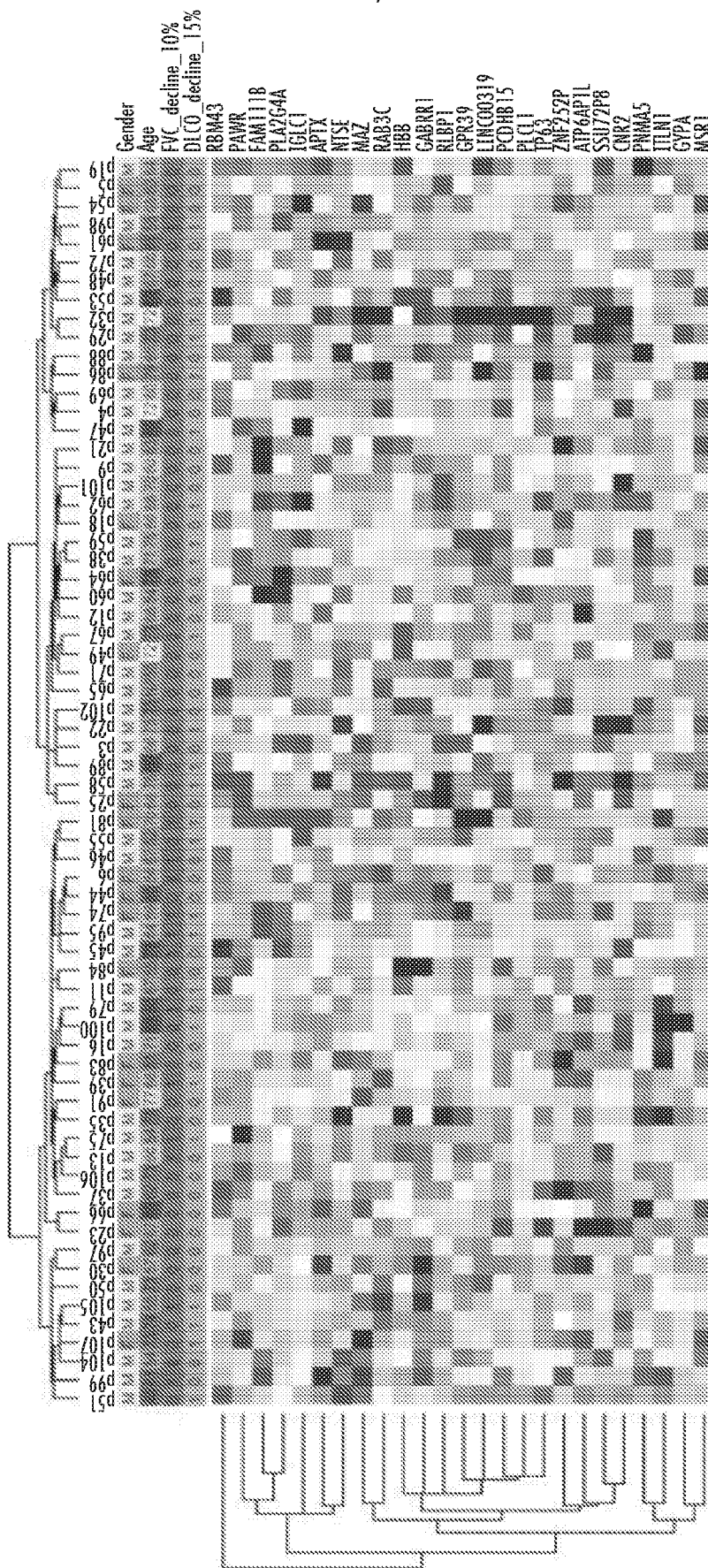


FIG. 5B

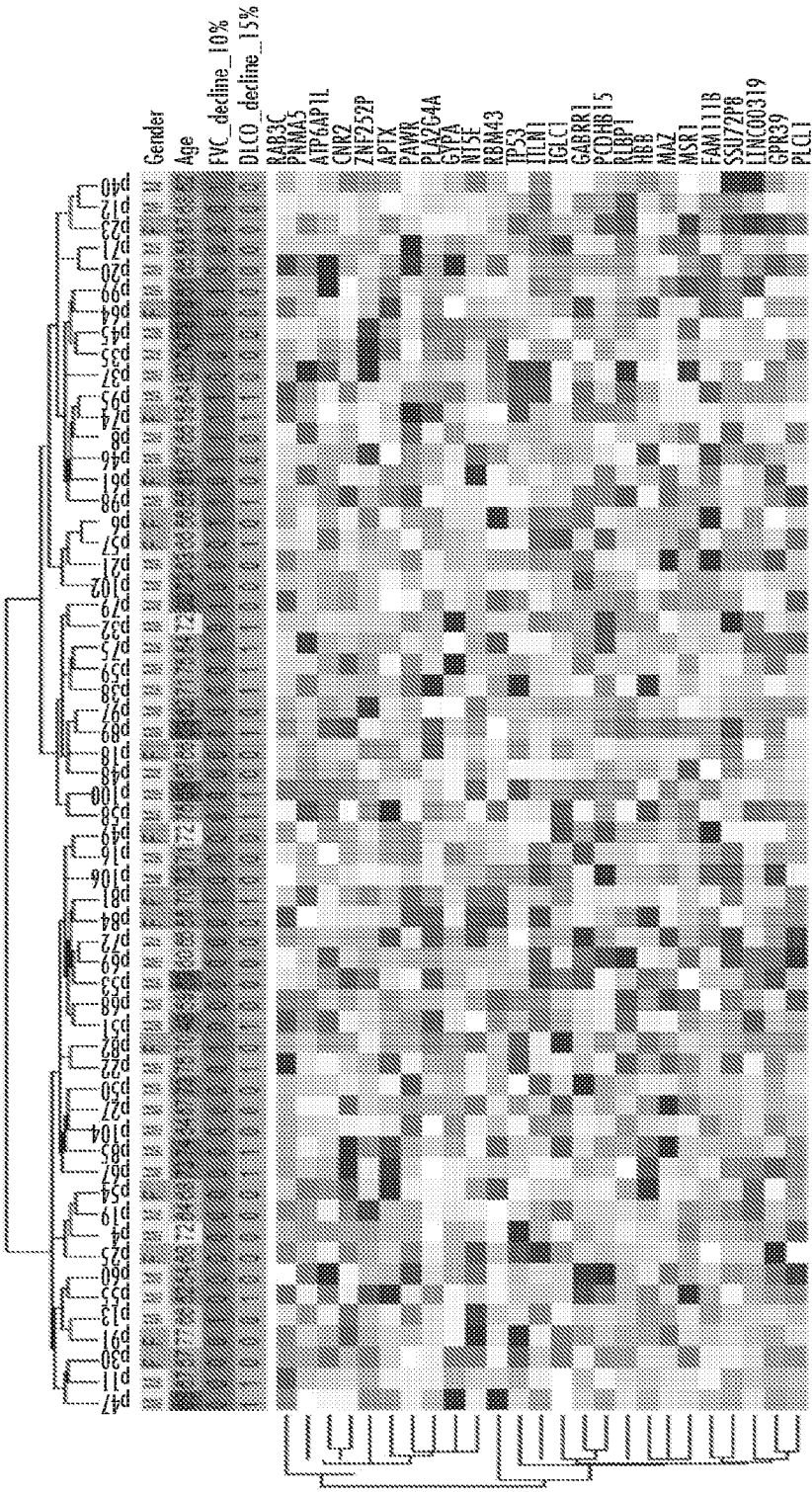


FIG. 5C

11/15

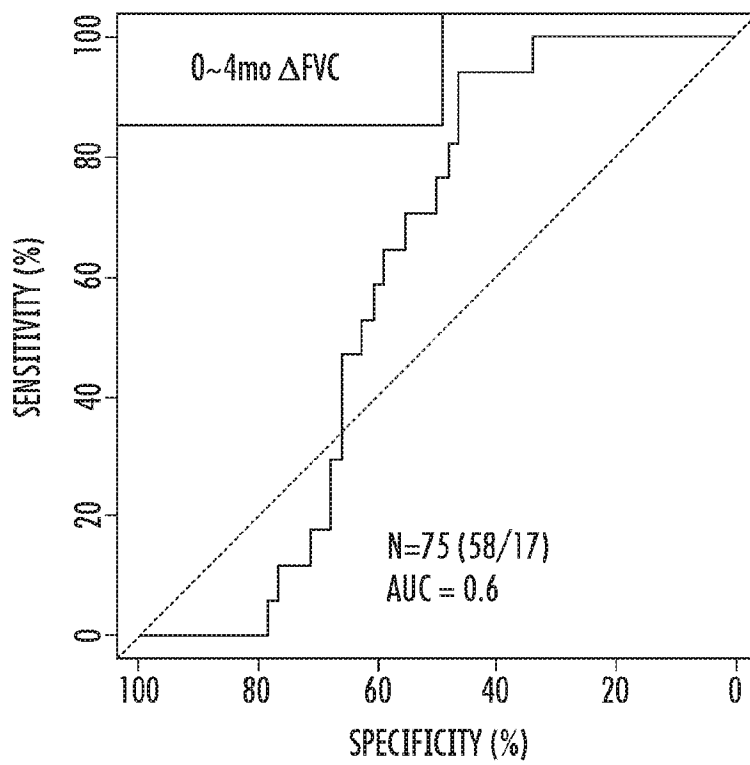


FIG. 6A

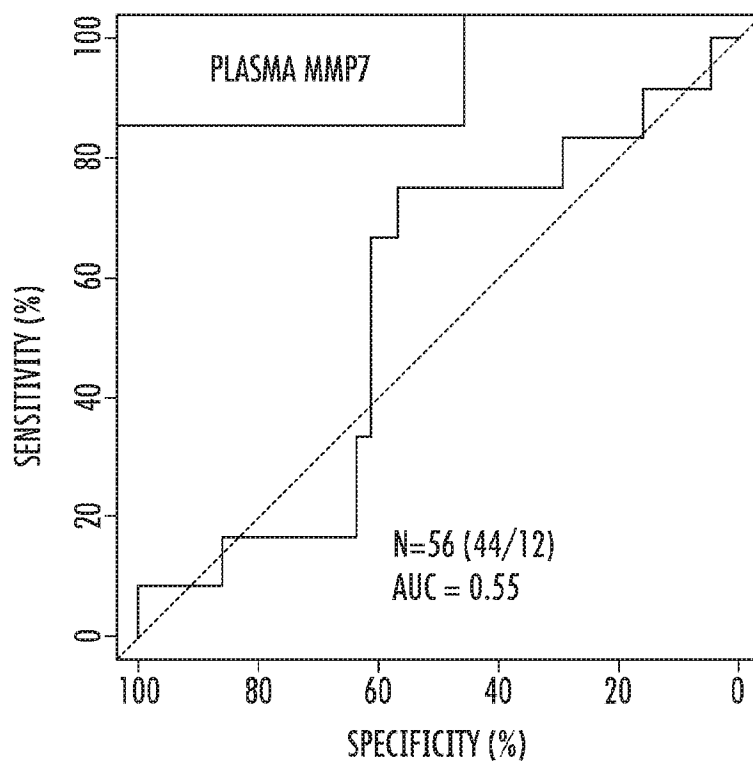


FIG. 6B

12/15

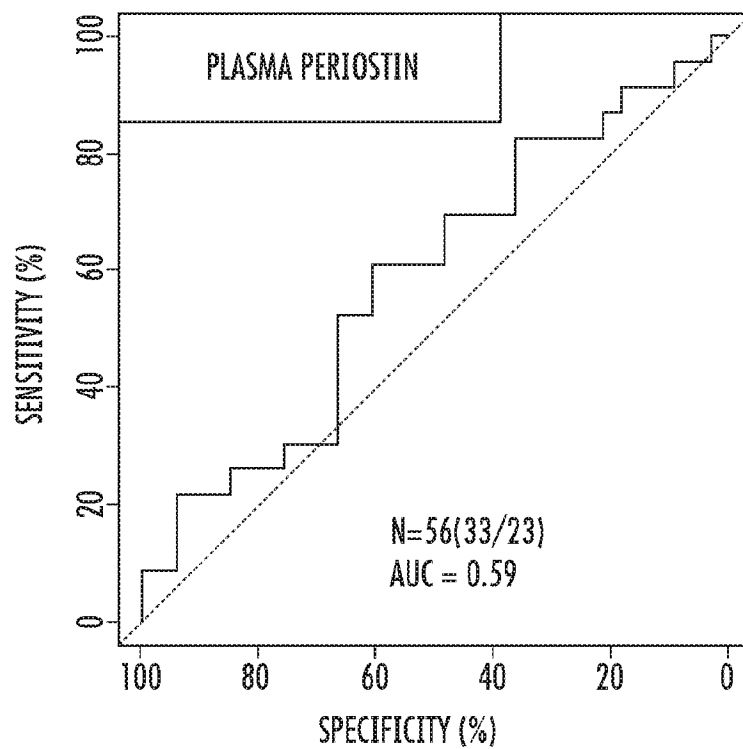


FIG. 6C

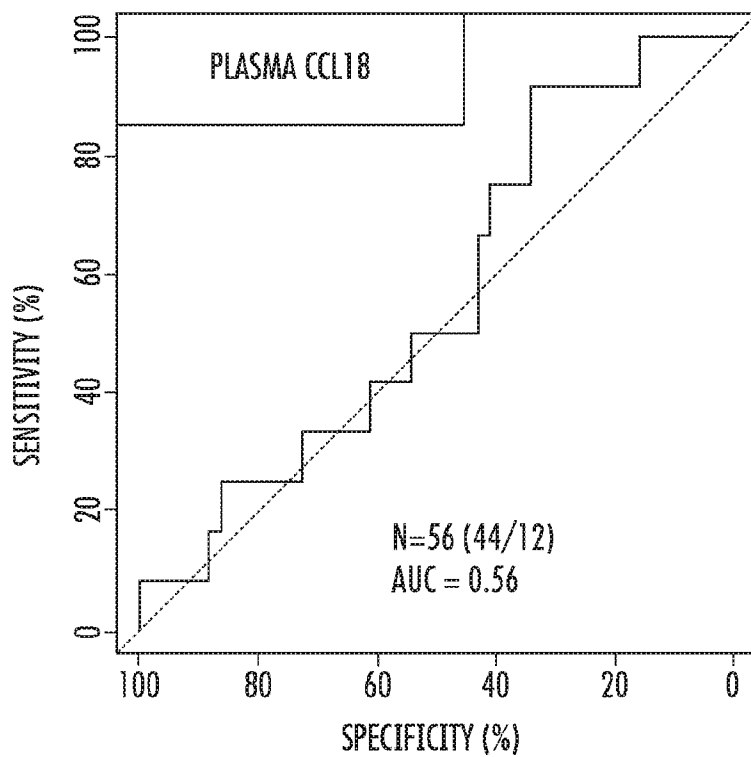


FIG. 6D

13/15

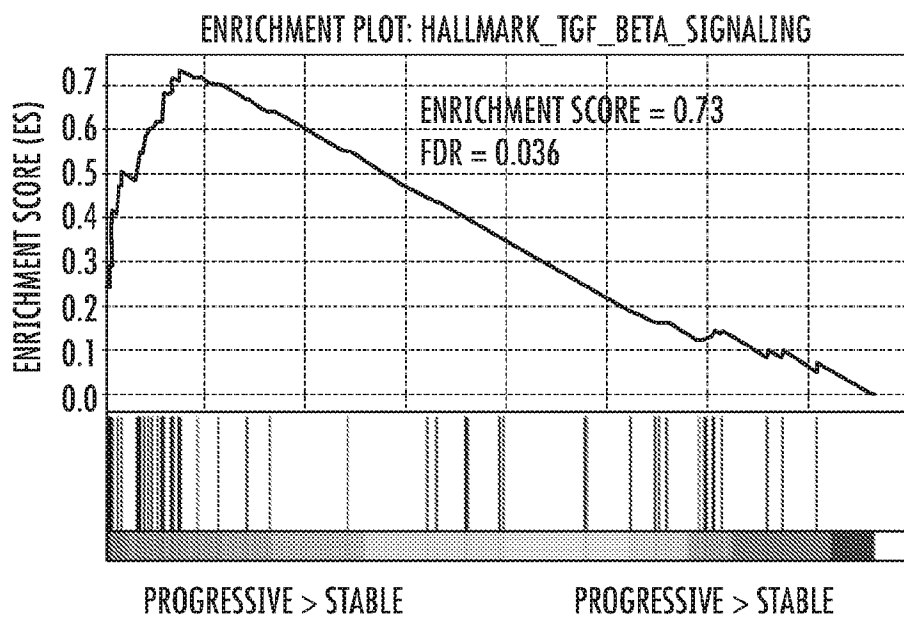


FIG. 7A

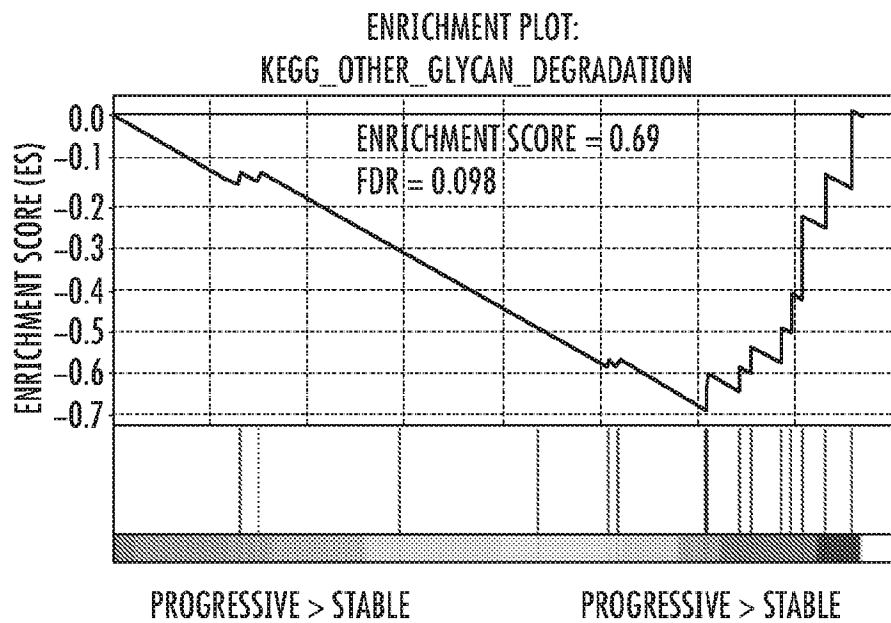


FIG. 7B

14/15

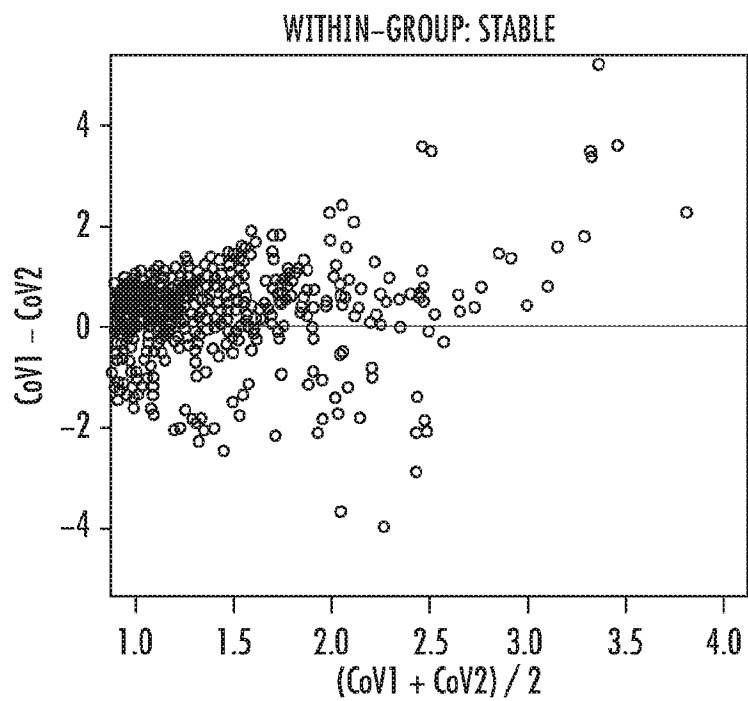


FIG. 8A

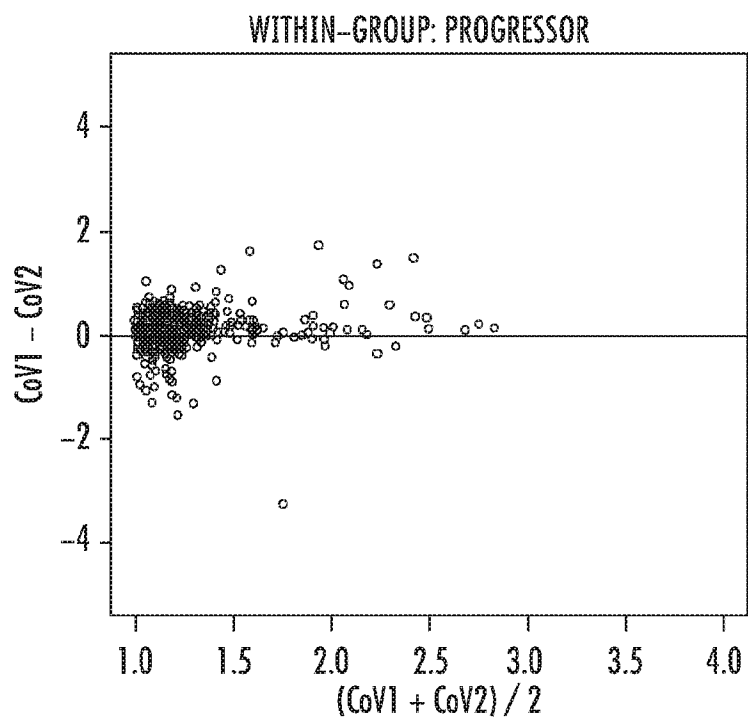
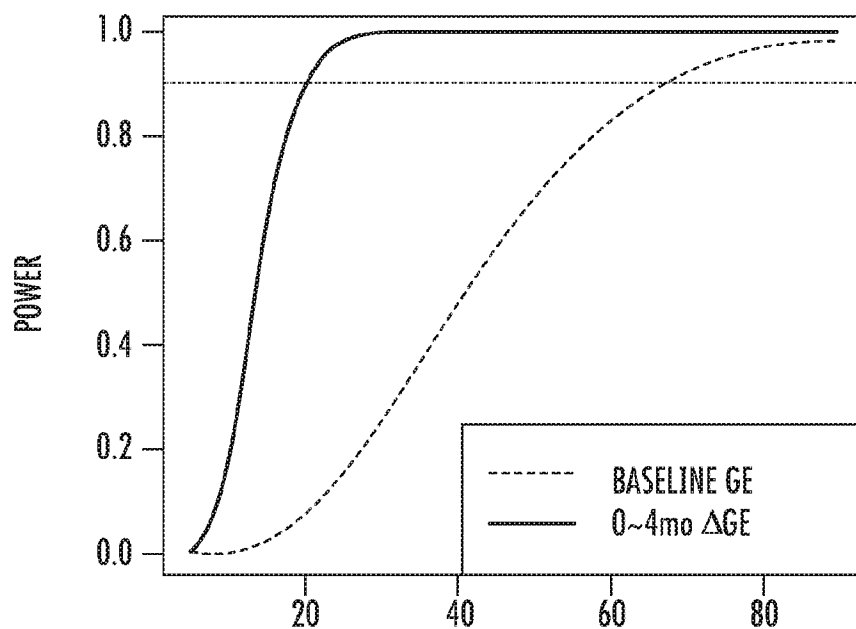


FIG. 8B

15/15



SAMPLE SIZE PER GROUP

FIG. 8C

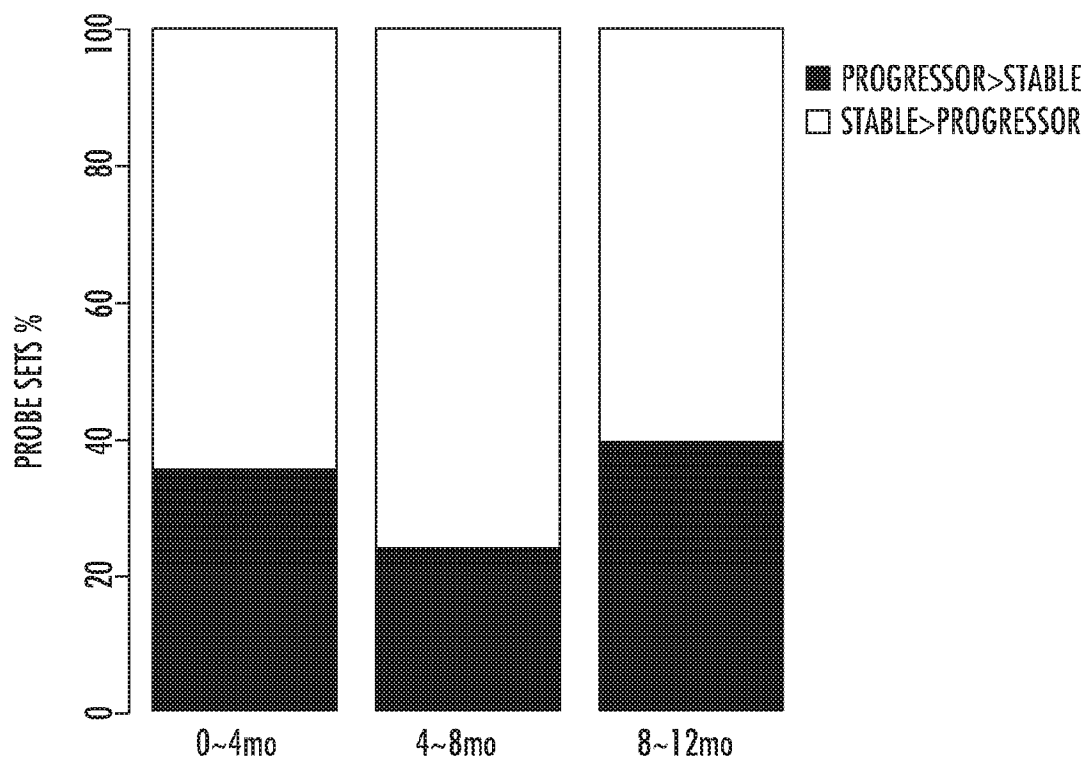


FIG. 8D

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2020/013367

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12P 19/34; C12Q 1/68; C12Q 1/6883; G01N 33/53; G01N 33/569; G01N 33/68 (2020.01)
 CPC - C12Q 1/6883; C12Q 2600/118; G01N 33/56966 (2020.02)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

USPC - 435/6.12; 435/7.1; 435/7.21; 506/9 (keyword delimited)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2013/0116166 A1 (DESAI et al) 09 May 2013 (09.05.2013) entire document	1, 10-13, 19-23, 29-33
Y	EP 2860525 A1 (ROCHE DIAGNOSTICS GMBH et al) 15 April 2015 (15.04.2015) entire document	1, 10-13, 19-23, 29-33
Y	US 2016/0024583 A1 (ALLEGRO DIAGNOSTICS CORP.) 28 January 2016 (28.01.2016) entire document	13, 23, 33
A	WO 2016/042114 A1 (F. HOFFMANN-LA ROCHE AG et al) 24 March 2016 (24.03.2016) entire document	1-3, 10-14, 19-23, 29-33
A	US 2018/0320234 A1 (THE UNIVERSITY OF CHICAGO et al) 08 November 2018 (08.11.2018) entire document	1-3, 10-14, 19-23, 29-33
A	US 2018/0101642 A1 (YALE UNIVERSITY) 12 April 2018 (12.04.2018) entire document	1-3, 10-14, 19-23, 29-33

☐ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

27 February 2020

Date of mailing of the international search report

01 APR 2020

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, VA 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Blaine R. Copenheaver

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2020/013367

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☒ Claims Nos.: 4-9, 15-18, 24-28, 34-39
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.