(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 August 2011 (04.08.2011)

PCT

(10) International Publication Number
WO 2011/094345 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01) *G01N 33/68* (2006.01)

(21) International Application Number:
PCT/US2011/022621

(22) International Filing Date:
26 January 2011 (26.01.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

61/298,473	26 January 2010 (26.01.2010)	US
61/298,814	27 January 2010 (27.01.2010)	US
61/323,238	12 April 2010 (12.04.2010)	US
61/323,760	13 April 2010 (13.04.2010)	US

(71) Applicant *(for all designated States except US)*: **NATIONAL JEWISH HEALTH** [US/US]; 1400 Jackson Street, Room M206b, Denver, Colorado 80206 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **SCHWARTZ, David** [US/US]; University of Colorado, Denver, 12631 East 17th Avenue, B178, Aurora, Colorado 80045 (US). **SEIBOLD, Max** [US/US]; 1400 Jackson Street, Room M206b, Denver, Colorado 80206 (US).

(74) Agents: **JOHNS, Carol P.** et al.; Kilpatrick Townsend and Stockton LLP, Two Embarcadero Center, 8th Floor, San Francisco, California 94111 (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, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2011/094345 A1

(54) Title: METHODS AND COMPOSITIONS FOR RISK PREDICTION, DIAGNOSIS, PROGNOSIS, AND TREATMENT OF PULMONARY DISORDERS

(57) Abstract: The invention provides diagnostic and therapeutic targets for pulmonary disease, in particular, fibrotic lung disease. The inventors have found that a genetic variant *MUC5B* gene is associated with increased expression of the gene, increased risk of developing a pulmonary disease, and an improved prognosis and survival among those developing the pulmonary disease.

METHODS AND COMPOSITIONS FOR RISK PREDICTION, DIAGNOSIS, PROGNOSIS, AND TREATMENT OF PULMONARY DISORDERS

CROSS-REFERENCES TO RELATED APPLICATIONS

[0001] This application claims priority to US Provisional Application No. 61/298,473, filed January 26, 2010, US Provisional Application No. 61/298,814, filed January 27, 2010, US Provisional Application No. 61/323,238, filed April 12, 2010, and US Provisional Application No. 61/323,760, filed April 13, 2010, the disclosures of which are incorporated herein in their entireties.

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

[0002] The present invention was supported at least in part by government funding from the NIH Intramural Research Program of the National Inst. of Environmental Health Sciences (Grant No. Z01-ES101947) and the National Heart, Lung, and Blood Inst. (Grant Nos. U01-HL067467, R01-HL095393, R01-HL097163, P01-HL092870, and RC2-HL101715). The government has certain rights in the invention.

BACKGROUND OF THE INVENTION

[0003] Pulmonary fibrosis disorders are a growing concern in human and non-human populations. Pulmonary fibrosis is associated with a number of complex disorders (*e.g.*, Herman-Pudlak Syndrome, tuberous sclerosis, neurofibromatosis, and dyskeratosis congenital). Idiopathic interstitial pneumonia (IIP) represents a class of chronic pulmonary fibrotic disorder characterized by progressive scarring of the alveolar interstitium leading to severe dyspnea, hypoxemia, and death. Idiopathic pulmonary fibrosis (IPF) is the most common type of IIP and currently has the highest mortality. Despite being an area of intensive research, the etiology of IPF is largely unknown. Familial clustering of IPF and differential susceptibility of individuals to fibrogenic dusts has implicated genetics in the development of this disorder. Genetic variants

in the telomerase reverse transcriptase (TERT), surfactant protein A1, and surfactant protein C genes have been implicated in development of familial interstitial pneumonia (FIP). However, these mutations only account for a small percentage of FIP cases. Familial association with IPF is 5-20%, and inheritance appears to be autosomal. The efficacy of current treatments, such as fibrogenic agents, is variable, indicating a need for more individualized treatment.

[0004] Mucins represent a family of glycoproteins associated with mucosal epithelia. Mucins can be associated with the cell membrane or secreted, and typically form a component of mucus. Abnormal expression or mutations in these proteins have been associated with adenocarcinomas, as well as pulmonary disorders such as asthma and bronchitis.

[0005] The present inventors have found that genetic variants of the *MUC5B* gene are associated with pulmonary disease, and can provide a useful tool for prognosing the course of disease and determining a course of treatment. In addition, the increased level of *MUC5B* expression that results from the disclosed genetic variants provides a novel therapeutic target for pulmonary diseases such as IIP, IPF, and FIP.

BRIEF SUMMARY OF THE INVENTION

[0006] Accordingly, in some embodiments, the invention provides methods and compositions for diagnosis, risk prediction, and determining the course of pulmonary disease. The invention further provides personalized methods of treatment for pulmonary diseases.

[0007] In some embodiments, the invention provides methods of determining whether a subject has or is at risk of developing a pulmonary disease, said method comprising determining (detecting) whether a subject expresses an elevated MUC5B RNA level or an elevated MUC5B protein level relative to a standard (*e.g.*, normal) control, wherein the presence of said elevated MUC5B RNA level or said elevated MUC5B protein level indicates said subject has or is at risk of developing a pulmonary disease. In some embodiments, the pulmonary disease is an interstitial lung disease, *e.g.*, a fibrotic interstitial lung disease, such as idiopathic pulmonary fibrosis or familial interstitial pneumonia.

[0008] The level of MUC5B RNA or protein can be determined using an *in vitro* assay or *in vivo* imaging assay. In some embodiments, said elevated MUC5B protein level or said elevated MUC5B RNA level is determined from a biological sample from the subject, *e.g.*, a pulmonary tissue or bodily fluid of said subject. The bodily fluid can be, *e.g.*, whole blood, plasma, serum,

urine, sputum, saliva, a bronchoalveolar lavage sample, or exhaled breath condensate. In some embodiments, the sample is further processed, *e.g.*, to separate cellular components or subcellular components. For example, the determining can further comprises separating cells from the remaining sample, or isolating exosomes or subcellular vesicles.

[0009] In some embodiments, the method further comprises administering a treatment to the subject, *e.g.*, a pulmonary disease treatment, or interstitial lung disease treatment. In some embodiments, the treatment is a mucolytic agent. In some embodiments, the treatment is a MUC5B antagonist. In some embodiments, the method further comprises determining a second MUC5B RNA level or MUC5B protein level after administering said treatment and comparing said second level to the level observed before administering said treatment.

[0010] In some embodiments, the expression level of at least one additional pulmonary disease marker is determined and compared to a standard control. For example, the at least one additional pulmonary disease marker can be selected from the group consisting of Surfactant Protein A, Surfactant Protein D, KL-6/MUC1, CC16, CK-19, Ca 19-9, SLX, MCP-1, MIP-1a, ITAC, glutathione, type III procollagen peptide, sIL-2R, ACE, neopterin, beta-glucuronidase, LDH, CCL-18, CCL-2, CXCL12, MMP7, and osteopontin. An aberrant expression level of the pulmonary disease marker indicates that the subject has or is at risk of developing a pulmonary disease. In some embodiments, the aberrant expression is elevated relative to a normal control. In some embodiments, the aberrant expression is reduced relative to a normal control. In some embodiments, the method comprises determining whether the genome of the subject comprises a genetic variant of the at least one additional pulmonary disease marker selected from the group consisting of Surfactant Protein A2, Surfactant Protein B, Surfactant Protein C, TERC, TERT, IL-1RN, IL-1α, IL-1β, TNF, Lymphotoxin α, TNF-RII, IL-10, IL-6, IL-12, IFNγ, TGFβ, CR1, ACE, IL-8, CXCR1, CXCR2, MUC1 (KL6), and MUC5AC, wherein the presence of a genetic variant of the at least one additional pulmonary disease marker is indicative that the subject has or is at risk of developing a pulmonary disease. In some embodiments, the method does not comprise determining whether the genome of the subject comprises a genetic variant of MUC5AC.

[0011] In some embodiments, the standard control is obtained from normal, non-diseased sample. In some embodiments, the standard control is from a different individual or pool of individuals. In some embodiments, the standard control is a standard obtained from a population

of individuals that do not have a pulmonary disease. In some embodiments, the standard control is obtained from the same individual, *e.g.*, obtained at a different time, *e.g.*, prior to exposure to an airway stressor. Typically, when detecting or determining the expression level of a given RNA or protein (*e.g.*, MUC5B), the same RNA or protein is detected in the standard control. However, in some embodiments, a different RNA or protein can be detected and the ratio used to determine whether the RNA or protein level from the subject is elevated. Moreover, in some embodiments, the method can comprise comparison to a positive control, *e.g.*, from a known pulmonary disease sample, or a sample from a known individual or pool of individuals that carry a genetic variant *MUC5B* gene or have elevated MUC5B expression.

[0012] In some embodiments, the invention provides methods of determining whether a subject has or is at risk of developing a pulmonary disease, said method comprising detecting (determining) whether a genome of a subject comprises a genetic variant *MUC5B* gene, wherein the presence of said genetic variant *MUC5B* gene indicates said subject has or is at risk of developing a pulmonary disease. In some embodiments, the pulmonary disease is an interstitial lung disease, *e.g.*, a fibrotic interstitial lung disease, such as idiopathic pulmonary fibrosis or familial interstitial pneumonia.

[0013] In some embodiments, the genetic variant *MUC5B* gene in said subject results in elevated expression of MUC5B RNA or MUC5B protein. In some embodiments, the subject is homozygous for said genetic variant *MUC5B* gene. In some embodiments, the subject is heterozygous for said genetic variant *MUC5B* gene. In some embodiments, the subject lacks the genetic variant *MUC5B* gene. In some embodiments, the genetic variant *MUC5B* gene is a genetic variant regulatory region *MUC5B* gene, *e.g.*, a genetic variant promoter *MUC5B* gene. In some embodiments, the genetic variant *MUC5B* gene has a single nucleotide polymorphism (SNP). In some embodiments, the SNP is selected from the group consisting of single nucleotide polymorphism is rs2672792, rs72636989, *MUC5B*-Prm1, rs2672794, rs35705950, *MUC5B*-Prm2, rs11042491, rs2735726, rs868902, *MUC5B*-Prm3, *MUC5B*-Prm4, *MUC5B*-Prm5, rs868903, *MUC5B*-Prm6, rs885455, rs885454, *MUC5B*-Prm7, rs7115457, rs7118568 rs56235854 and rs2735738. In some embodiments, the presence of more than one SNP is determined. In some embodiments, the SNP is rs35705950.

[0014] In some embodiments, the genetic variant *MUC5B* gene comprises a first single nucleotide polymorphism (SNP) and a second SNP. In some embodiments, the first SNP is

present within a first DNA strand and said second SNP is present within a second DNA strand. In some embodiments, the first and second SNP are present within the same DNA strand.

[0015] In some embodiments, the determining comprises use of at least one sequence selected from the group consisting of SEQ ID NOs:20-53 to determine whether the genome of the subject comprises a genetic variant *MUC5B* gene, *e.g.*, by using an appropriate nucleic acid assay to detect the variant nucleotide in the selected sequence. For example, the determining can comprise use of one or more of the sequences of SEQ ID NOs:20-53 in an RT-PCR, array hybridization, or other appropriate SNP detection method as described herein. In some embodiments, the determining comprises (i) contacting a sample from the subject with a nucleic acid probe having at least 10 contiguous nucleotides of at least one of the sequences selected from SEQ ID NOs:20-53, or its complement, wherein said 10 contiguous nucleotides span the genetic variant nucleotide (*i.e.*, the position of the SNP shown for each sequence), and (ii) determining whether the nucleic acid probe hybridizes to a nucleic acid in the sample. In some embodiments, the at least one sequence includes SEQ ID NO:24, wherein the presence of a T at position 28 of SEQ ID NO:24 indicates a genetic variant *MUC5B* gene, and that the subject has or will have an attenuated form of the pulmonary disease. The presence of a G at position 28 of SEQ ID NO:24 indicates that the subject has or will have a more severe form of the pulmonary disease (*e.g.*, where the subject is homozygous for G at position 28, or lacking a genetic variant promoter *MUC5B* gene).

[0016] In some embodiments, the method further comprises determining whether said individual expresses an elevated MUC5B RNA level or an elevated MUC5B protein level relative to a standard control, wherein the presence of said elevated MUC5B RNA level or said elevated MUC5B protein level further indicates said subject has or is at risk of developing a pulmonary disease. Said step of determining can be carried out as discussed above.

[0017] In some embodiments, the method does not comprise determining whether the individual expresses an elevated level of MUC5AC RNA or protein. In some embodiments, the method does not comprise determining whether said individual expresses an elevated level of a second RNA or protein other than a MUC5B RNA or protein. In some embodiments, the method does not comprise determining whether said individual expresses an elevated level of a second RNA or protein other than a MUC5B RNA or protein, unless said second RNA or protein is a MUC5AC RNA or protein.

[0018] In some embodiments, the method further comprises administering a treatment to the subject, *e.g.*, a pulmonary disease treatment, or interstitial lung disease treatment. In some embodiments, the treatment is a mucolytic agent. In some embodiments, the treatment is a MUC5B antagonist, *e.g.*, small molecule that inhibits MUC5B production or activity. In some embodiments, the method further comprises determining a second MUC5B RNA level or MUC5B protein level after administering said treatment and comparing said second level to the level observed before administering said treatment.

[0019] In some embodiments, the method further comprises determining whether the genome of the subject comprises at least one additional genetic variant pulmonary disease marker gene. In some embodiments, the at least one additional pulmonary disease marker can be selected from the group consisting of Surfactant Protein A2, Surfactant Protein B, Surfactant Protein C, TERC, TERT, IL-1RN, IL-1α, IL-1β, TNF, Lymphotoxin α, TNF-RII, IL-10, IL-6, IL-12, IFNγ, TGFβ, CR1, ACE, IL-8, CXCR1, CXCR2, MUC1 (KL6), or MUC5AC. The presence of an additional genetic variant pulmonary disease marker gene can indicate that the subject is at risk of or has a pulmonary disease.

[0020] In some embodiments, the presence of the genetic variant *MUC5B* gene indicates that the subject has an attenuated form of the pulmonary disease. That is, the subject will have a reduced severity of symptoms, more gradual loss of lung function, or increased survival compared to the normal, non-attenuated form of the pulmonary disease, *i.e.*, compared to the pulmonary disease as it occurs in an individual that does not have a genetic variant *MUC5B* gene.

[0021] Thus, in some embodiments, the invention provides methods of prognosing a pulmonary disease in a patient, said method comprising determining whether a genome of a subject comprises a genetic variant *MUC5B* gene, wherein the presence of said genetic variant *MUC5B* gene indicates an attenuated form of said pulmonary disease in said patient relative to the absence of said genetic variant *MUC5B* gene. The absence of a genetic variant *MUC5B* gene can indicate that the patient has a more aggressive form of said pulmonary disease. In some embodiments, the pulmonary disease is an interstitial lung disease, *e.g.*, a fibrotic interstitial lung disease, such as idiopathic pulmonary fibrosis or familial interstitial pneumonia. Said genetic variant *MUC5B* gene can be as described above.

[0022] In some embodiments, the method further comprises setting a course of treatment for the subject, *e.g.*, based on the presence of a genetic variant *MUC5B* gene in the subject. For example, the presence of a genetic variant *MUC5B* gene, or the level of *MUC5B* gene expression, can be determined in the subject, a treatment administered to the subject, and the progress of the subject monitored, *e.g.*, by monitoring *MUC5B* expression over time or other pulmonary diagnostic indicators, and determining whether further treatment is necessary. Thus, in some embodiments, the method further comprises administering pulmonary disease treatment or an interstitial lung disease treatment to the subject. In some embodiments, the method further comprises determining whether the genome of the subject comprises a genetic variant *MUC5B* gene, wherein the presence of a genetic variant *MUC5B* gene indicates an attenuated form of said interstitial lung disease in said subject.

[0023] In some embodiments, the invention provides methods of determining whether a pulmonary disease is progressing in pulmonary disease patient, said method comprising: (i) determining a first level of MUC5B RNA or first level of MUC5B protein in said patient at a first time point; (ii) determining a second level of MUC5B RNA or second level of MUC5B protein in said patient at a second time point; and (iii) comparing the second level of MUC5B RNA to the first level of MUC5B RNA or comparing the second level of MUC5B protein to the first level of MUC5B protein, wherein if the second level of MUC5B RNA is greater than the first level of MUC5B RNA or if the first level of MUC5B protein is greater than the first level of MUC5B protein, the pulmonary disease is progressing in the patient. In some embodiments, the pulmonary disease is an interstitial lung disease, *e.g.*, a fibrotic interstitial lung disease, such as idiopathic pulmonary fibrosis or familial interstitial pneumonia.

[0024] In some embodiments, the method further comprises determining the rate of progression based on said comparing. That is, an rapid increase in MUC5B expression in a short time is correlated with more rapid progression of the pulmonary disease. In some embodiments, said determining said first level of MUC5B RNA or first level of MUC5B protein and said second level of MUC5B RNA or second level of MUC5B protein comprises normalizing said first level of MUC5B RNA or first level of MUC5B protein and said second level of MUC5B RNA or second level of MUC5B protein to a level of RNA or protein expressed from a standard gene in said interstitial lung disease patient, *e.g.*, GAPDH, beta-actin, HPRT1, beta-tubulin, or beta-20 microglobulin.

[0025] In some embodiments, the invention provides methods of treating, preventing, or ameliorating a pulmonary disease in a subject in need thereof, the method comprising administering to said patient an effective amount of a MUC5B antagonist, wherein said antagonist reduces the expression of the *MUC5B* gene or reduces the activity of the MUC5B protein as compared to the expression or activity in the absence of said MUC5B antagonist, thereby treating, preventing, or ameliorating the pulmonary disease in the subject. In some embodiments, the MUC5B antagonist is a nucleic acid, *e.g.*, a pRNA, siRNA, or antisense sequence, and reduces expression of the *MUC5B* gene. In some embodiments, the MUC5B antagonist is a small molecule, *e.g.*, that reduces translation of MUC5B mRNA or packaging or activity of the MUC5B protein. In some embodiments, the MUC5B antagonist is selected from the group consisting of: a MUC5B antibody or MUC5B-binding fragment thereof, a MUC5B-binding aptamer, and a mucolytic agent. In some embodiments, the MUC5B antagonist nucleic acid is capable of hybridizing to at least a 10-nucleotide contiguous sequence of a MUC5B encoding target nucleic acid sequence. In some embodiments, the method further comprises monitoring the subject, *e.g.*, by determining the level of MUC5B RNA or protein before and after said administering, or at one or more time points after said administering. Thus, in some embodiments, the method of treatment includes a step of determining whether the genome of the subject comprises a genetic variant *MUC5B* gene, and/ or a step of determining whether the subject has an elevated level of MUC5B RNA or protein, as described herein.

[0026] In some embodiments, the invention provides methods of identifying a candidate pulmonary disease treatment compound, said method comprising: (i) contacting a test compound with a MUC5B protein; (ii) allowing said test compound to inhibit the activity of said MUC5B protein; and (iii) selecting the test compound that inhibits the activity of said MUC5B protein, thereby identifying a candidate pulmonary disease treatment compound. In some embodiments, the method is carried out *in vivo*, *e.g.*, in an animal model for pulmonary disease. In some embodiments, the method is carried out *in vitro*.

[0027] In some embodiments, the invention provides methods of identifying a candidate pulmonary disease treatment compound, said method comprising: (i) contacting a test compound with a MUC5B secreting cell; (ii) allowing said test compound to inhibit secretion of MUC5B protein from said MUC5B secreting cell; and (iii) selecting the test compound that inhibits secretion of MUC5B protein from said MUC5B secreting cell, thereby identifying a candidate pulmonary disease treatment compound. In some embodiments, said MUC5B secreting cell is in

vitro. In some embodiments, said MUC5B secreting cell forms part of a pulmonary tissue. In some embodiments, said pulmonary tissue forms part of an organism, *i.e.*, the method is carried out *in vivo*. In some embodiments, the organism is a mammal, *e.g.*, an animal model or a human.

[0028] The invention further provides kits, *e.g.*, for determining whether a subject expresses an elevated level of MUC5B RNA or MUC5B protein, or carries a genetic variant *MUC5B* gene. In some embodiments, the kit comprises (a) a MUC5B binding agent capable of binding to a substance selected from the group consisting of (i) a genetic variant MUC5B gene sequence; (ii) a MUC5B RNA or fragment thereof; and (iii) a MUC5B protein or fragment thereof, and (b) a detecting reagent or a detecting apparatus capable of indicating binding of said MUC5B binding agent to said substance. In some embodiments, the MUC5B binding agent is labeled, *e.g.*, with a fluorescent label or radioisotope. In some embodiments, the kit further comprises a sample collection device for collecting a sample from the subject. In some embodiments, the MUC5B binding agent binds a genetic variant *MUC5B* gene in the promoter region. In some embodiments, the kit further comprises at least one control sample, *e.g.*, a non-variant *MUC5B* gene sequence or a sample from a normal, non-disease control.

BRIEF DESCRIPTION OF THE DRAWINGS

[0029] Figure 1 represents a flow chart related to the genetic study design described herein.

[0030] Figure 2 represents a Multipoint LOD score graphs for whole genome screen (884 markers with an average inter-marker distance of 4.2 centimorgans (CM)) in 82 families with two or more cases of IIP.

[0031] Figure 3 illustrates pair-wise linkage disequilibrium (LD) plot for SNPs significantly associated with IPF or FIP by allelic association test in genetic screen of lung-expressed gel-forming mucins. LD values displayed are calculated by the r2 statistic for the mucin genetic screen IPF subjects (n=492). Multi-colored graphic about the plot indicates the approximate location of these SNPs within the gel-forming mucin region. The highly significant *MUC5B* promoter SNP (rs35705950) and the corresponding pairwise LD values are highlighted in red. Intergenic region is abbreviated as Int, and the *MUC5B* Promoter is abbreviated as Pr. LD patterns were qualitatively similar in the controls although in most instances the LD was weaker among controls.

[0032] Figures 4A-4C represent illustrations of *MUC5B* gene expression in IPF (N=33) and unaffected subjects (N=47) stratified by *MUC5B* promoter SNP (rs35705950) genotype and smoking status. A. *MUC5B* gene expression among unaffected and IPF subjects colored coded based on whether subjects are wildtype (dark grey) or heterozygous for the *MUC5B* promoter SNP (light grey). B. Comparison of *MUC5B* expression in unaffected subjects, among unaffected smokers only, and among unaffected non-smokers only, by *MUC5B* promoter SNP genotype. C. Comparison of *MUC5B* expression in all IPF subjects, among IPF smokers only, and among IPF non-smokers only, by *MUC5B* promoter SNP genotype. Lines represent group medians and the expression of *MUC5B* is determined relative to GAPDH expression

[0033] Figures 5A-5C represent MUC5B immunohistochemistry of unaffected and IPF tissue. Tissue sections stained for MUC5B distribution in both the unaffected and IPF lung show strong specific cytoplasmic staining within secretory columnar cells of the bronchi and larger proximal bronchioles (Fig. 5A). In subjects with IPF, regions of dense accumulation of MUC5B were observed in areas of microscopic honeycombing and involved patchy staining of the metaplastic epithelia lining the honeycomb cysts (Fig. 5B), as well as the mucus plugs within the cysts (Fig. 5C).

DETAILED DESCRIPTION OF THE INVENTION

[0034] The invention provides novel methods and compositions for diagnosing and predicting the severity of pulmonary disease, and a novel therapeutic target for ameliorating pulmonary disease. The inventors have found that individuals carrying genetic variants of the *MUC5B* gene that have elevated expression of the gene have an increased likelihood of developing a pulmonary disease, *e.g.*, an interstitial lung disease such as fibrotic interstitial lung disease, idiopathic pulmonary fibrosis, familial interstitial pneumonia, etc. The presence of some genetic variations in the *MUC5B* gene, while increasing the likelihood of a pulmonary disease, are indicative of an attenuated form of the disease, *e.g.*, a more gradual progression of symptoms and improved survival.

I. Definitions

[0035] The terms "pulmonary disease," "pulmonary disorder," "lung disease," etc. are used interchangeably herein. The term is used to broadly refer to lung disorders characterized by difficulty breathing, coughing, airway discomfort and inflammation, increased mucus, and/or pulmonary fibrosis.

[0036] Mucins are a family of high molecular weight, heavily glycosylated proteins (glycoproteins) produced by mammalian epithelia. Secreted, gel-forming mucins form a component of mucus. Typically, the N- and C- terminal ends of mucin proteins are lightly glycosylated, but rich in di-sulfide bond-forming cysteine residues.

[0037] Mucin 5b (MUC5B) is a gel-forming mucin expressed in airway epithelial tissue. Additional gel-forming mucins, MUC2, MUC5AC, and MUC6, have been mapped to the same chromosomal region on human chromosome 11. MUC5B is further characterized in Desseyn *et al.* (1996) *J. Biol. Chem.* 273:30157-64.

[0038] The term "genetic variant," in the context of a particular gene, refers a gene with a variant (*e.g.*, non-standard or abnormal) nucleic acid sequence. The gene includes coding and non-coding sequences, such as regulatory regions. Genetic variants include mutations and polymorphic sequences. Thus, the genetic variant may affect the expression or activity of the gene or gene product. The genetic variant may be an insertion of one or more nucleotides, deletion of one or more nucleotides, or a substitution of one or more nucleotides. A single nucleotide polymorphism (SNP) is an example of a genetic variant.

[0039] The term "genetic variant *MUC5B* gene" refers to a *MUC5B* genetic variant (a *MUC5B* gene with a genetic variation as described above). The term "genetic variant promoter *MUC5B* gene" refers to a variation that is specifically in the promoter region of the *MUC5B* gene. Similarly, "genetic variant regulatory region *MUC5B* gene" and "genetic variant intronic *MUC5B* gene" localize the variation within the *MUC5B* gene. An example of a genetic variant *MUC5B* gene is rs35705950, which includes a SNP in the promoter region.

[0040] An "airway mucosal sample" can be obtained using methods known in the art, *e.g.*, a bronchial epithelial brush as described herein. Additional methods include endobronchial biopsy, bronchial wash, bronchoalveolar lavage, whole lung lavage, transendoscopic biopsy, and transtracheal wash.

[0041] The terms "subject," "patient," "individual," etc. are not intended to be limiting and can be generally interchanged. That is, an individual described as a "patient" does not necessarily have a given disease, but may be merely seeking medical advice.

[0042] A "control" sample or value refers to a sample that serves as a reference, usually a known reference, for comparison to a test sample. For example, a test sample can be taken from

a patient suspected of having a given pulmonary disease and compared to samples from a known pulmonary disease patient, known genetic variant *MUC5B* carrier, or a known normal (non-disease) individual. A control can also represent an average value gathered from a population of similar individuals, *e.g.*, pulmonary disease patients or healthy individuals with a similar medical background, same age, weight, etc. A control value can also be obtained from the same individual, *e.g.*, from an earlier-obtained sample, prior to disease, or prior to treatment. One of skill will recognize that controls can be designed for assessment of any number of parameters.

[0043] One of skill in the art will understand which controls are valuable in a given situation and be able to analyze data based on comparisons to control values. Controls are also valuable for determining the significance of data. For example, if values for a given parameter are widely variant in controls, variation in test samples will not be considered as significant.

[0044] As used herein, the terms "pharmaceutically" acceptable is used synonymously with physiologically acceptable and pharmacologically acceptable. A pharmaceutical composition will generally comprise agents for buffering and preservation in storage, and can include buffers and carriers for appropriate delivery, depending on the route of administration.

[0045] The terms "dose" and "dosage" are used interchangeably herein. A dose refers to the amount of active ingredient given to an individual at each administration. For the present invention, the dose will generally refer to the amount of pulmonary disease treatment, anti-inflammatory agent, or MUC5B antagonist. The dose will vary depending on a number of factors, including the range of normal doses for a given therapy, frequency of administration; size and tolerance of the individual; severity of the condition; risk of side effects; and the route of administration. One of skill will recognize that the dose can be modified depending on the above factors or based on therapeutic progress. The term "dosage form" refers to the particular format of the pharmaceutical, and depends on the route of administration. For example, a dosage form can be in a liquid form for nebulization, *e.g.*, for inhalants, in a tablet or liquid, *e.g.*, for oral delivery, or a saline solution, *e.g.*, for injection.

[0046] As used herein, the terms "treat" and "prevent" are not intended to be absolute terms. Treatment can refer to any delay in onset, reduction in the frequency or severity of symptoms, amelioration of symptoms, improvement in patient comfort and/or respiratory function, etc. The effect of treatment can be compared to an individual or pool of individuals not receiving a given treatment, or to the same patient prior to, or after cessation of, treatment.

[0047] The term "prevent" refers to a decrease in the occurrence of pulmonary disease symptoms in a patient. As indicated above, the prevention may be complete (no detectable symptoms) or partial, such that fewer symptoms are observed than would likely occur absent treatment.

[0048] The term "therapeutically effective amount," as used herein, refers to that amount of the therapeutic agent sufficient to ameliorate the disorder, as described above. For example, for the given parameter, a therapeutically effective amount will show an increase or decrease of at least 5%, 10%, 15%, 20%, 25%, 40%, 50%, 60%, 75%, 80%, 90%, or at least 100%. Therapeutic efficacy can also be expressed as "-fold" increase or decrease. For example, a therapeutically effective amount can have at least a 1.2-fold, 1.5-fold, 2-fold, 5-fold, or more effect over a control.

[0049] The term "diagnosis" refers to a relative probability that a pulmonary disease is present in the subject. Similarly, the term "prognosis" refers to a relative probability that a certain future outcome may occur in the subject. For example, in the context of the present invention, prognosis can refer to the likelihood that an individual will develop a pulmonary disease, or the likely severity of the disease (*e.g.*, severity of symptoms, rate of functional decline, survival, etc.). The terms are not intended to be absolute, as will be appreciated by any one of skill in the field of medical diagnostics.

[0050] The terms "correlating" and "associated," in reference to determination of a pulmonary disease risk factor, refers to comparing the presence or amount of the risk factor (*e.g.*, dysregulation or genetic variation in a mucin gene) in an individual to its presence or amount in persons known to suffer from, or known to be at risk of, the pulmonary disease, or in persons known to be free of pulmonary disease, and assigning an increased or decreased probability of having/ developing the pulmonary disease to an individual based on the assay result(s).

[0051] "Nucleic acid" or "oligonucleotide" or "polynucleotide" or grammatical equivalents used herein means at least two nucleotides covalently linked together. Oligonucleotides are typically from about 5, 6, 7, 8, 9, 10, 12, 15, 25, 30, 40, 50 or more nucleotides in length, up to about 100 nucleotides in length. Nucleic acids and polynucleotides are a polymers of any length, including longer lengths, *e.g.*, 200, 300, 500, 1000, 2000, 3000, 5000, 7000, 10,000, etc. A nucleic acid of the present invention will generally contain phosphodiester bonds, although in some cases, nucleic acid analogs are included that may have alternate backbones, comprising,

e.g., phosphoramidate, phosphorothioate, phosphorodithioate, or O-methylphophoroamidite linkages (see Eckstein, Oligonucleotides and Analogues: A Practical Approach, Oxford University Press); and peptide nucleic acid backbones and linkages. Other analog nucleic acids include those with positive backbones; non-ionic backbones, and non-ribose backbones, including those described in U.S. Patent Nos. 5,235,033 and 5,034,506, and Chapters 6 and 7, ASC Symposium Series 580, *Carbohydrate Modifications in Antisense Research*, Sanghui & Cook, eds.. Nucleic acids containing one or more carbocyclic sugars are also included within one definition of nucleic acids. Modifications of the ribose-phosphate backbone may be done for a variety of reasons, *e.g.*, to increase the stability and half-life of such molecules in physiological environments or as probes on a biochip. Mixtures of naturally occurring nucleic acids and analogs can be made; alternatively, mixtures of different nucleic acid analogs, and mixtures of naturally occurring nucleic acids and analogs may be made.

[0052] The terms "identical" or percent "identity," in the context of two or more nucleic acids (*e.g.*, genomic sequences or subsequences, such as shown in SEQ ID NOs:20-53, or coding sequences) or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same (*i.e.*, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more identity over a specified region), when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. Such sequences are then said to be "substantially identical." This definition also refers to the compliment of a test sequence. Optionally, the identity exists over a region that is at least about 10 to about 100, about 20 to about 75, about 30 to about 50 amino acids or nucleotides in length.

[0053] An example of algorithms suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al., Nuc. Acids Res.* 25:3389-3402 (1977) and Altschul *et al., J. Mol. Biol.* 215:403-410 (1990), respectively. As will be appreciated by one of skill in the art, the software for performing BLAST analyses is publicly available through the website of the National Center for Biotechnology Information (ncbi.nlm.nih.gov).

[0054] The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one

or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers, those containing modified residues, and non-naturally occurring amino acid polymer.

[0055] The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function similarly to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, *e.g.*, hydroxyproline, γ-carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, *e.g.*, an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, *e.g.*, homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs may have modified R groups (*e.g.*, norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions similarly to a naturally occurring amino acid.

[0056] Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

[0057] "Conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical or associated, *e.g.*, naturally contiguous, sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode most proteins. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to another of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes silent variations of the nucleic acid. One of skill will recognize that in certain contexts each codon in a nucleic

acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, often silent variations of a nucleic acid which encodes a polypeptide is implicit in a described sequence with respect to the expression product, but not with respect to actual probe sequences.

[0058] As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the invention.typically conservative substitutions for one another: 1) Alanine (A), Glycine (G); 2) Aspartic acid (D), Glutamic acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W); 7) Serine (S), Threonine (T); and 8) Cysteine (C), Methionine (M) (*see, e.g.,* Creighton, *Proteins* (1984)).

[0059] A "label" or a "detectable moiety" is a composition detectable by spectroscopic, photochemical, biochemical, immunochemical, chemical, or other physical means. For example, useful labels include ^{32}P, fluorescent dyes, electron-dense reagents, enzymes (*e.g.*, as commonly used in an ELISA), biotin, digoxigenin, or haptens and proteins or other entities which can be made detectable, *e.g.*, by incorporating a radiolabel into a peptide or antibody specifically reactive with a target peptide. Any method known in the art for conjugating an antibody to the label may be employed, *e.g.,* using methods described in Hermanson, Bioconjugate Techniques 1996, Academic Press, Inc., San Diego.

[0060] A "labeled nucleic acid probe or oligonucleotide" is one that is bound, either covalently, through a linker or a chemical bond, or noncovalently, through ionic, van der Waals, electrostatic, or hydrogen bonds to a label such that the presence of the probe may be detected by detecting the presence of the label bound to the probe. Alternatively, method using high affinity interactions may achieve the same results where one of a pair of binding partners binds to the other, *e.g.*, biotin, streptavidin.

[0061] The phrase "selectively (or specifically) hybridizes to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence with a higher affinity, *e.g.*, under more stringent conditions, than to other nucleotide sequences (*e.g.*, total cellular or library DNA or RNA). One of skill in the art will appreciate that specific hybridization between nucleotides usually relies on Watson-Crick pair bonding between complementary nucleotide sequences.

[0062] The term "probe" or "primer", as used herein, is defined to be one or more nucleic acid fragments whose specific hybridization to a sample can be detected. A probe or primer can be of any length depending on the particular technique it will be used for. For example, PCR primers are generally between 10 and 40 nucleotides in length, while nucleic acid probes for, *e.g.*, a Southern blot, can be more than a hundred nucleotides in length. The probe may be unlabeled or labeled as described below so that its binding to the target or sample can be detected. The probe can be produced from a source of nucleic acids from one or more particular (preselected) portions of a chromosome, *e.g.*, one or more clones, an isolated whole chromosome or chromosome fragment, or a collection of polymerase chain reaction (PCR) amplification products. The length and complexity of the nucleic acid fixed onto the target element is not critical to the invention. One of skill can adjust these factors to provide optimum hybridization and signal production for a given hybridization procedure, and to provide the required resolution among different genes or genomic locations.

[0063] The probe may also be isolated nucleic acids immobilized on a solid surface (*e.g.*, nitrocellulose, glass, quartz, fused silica slides), as in an array. In some embodiments, the probe may be a member of an array of nucleic acids as described, for instance, in WO 96/17958. Techniques capable of producing high density arrays can also be used for this purpose (*see*, *e.g.*, Fodor (1991) *Science* 767-773; Johnston (1998) *Curr. Biol.* 8: R171-R174; Schummer (1997) *Biotechniques* 23: 1087-1092; Kern (1997) *Biotechniques* 23: 120-124; U.S. Patent No. 5,143,854). One of skill will recognize that the precise sequence of the particular probes described herein can be modified to a certain degree to produce probes that are "substantially identical" to the disclosed probes, but retain the ability to specifically bind to (*i.e.*, hybridize specifically to) the same targets or samples as the probe from which they were derived. Such modifications are specifically covered by reference to the individual probes described herein.

[0064] "Antibody" refers to a polypeptide comprising a framework region from an immunoglobulin gene or fragments thereof that specifically binds and recognizes an antigen, *e.g.*, a specific bacterial antigen. Typically, the "variable region" contains the antigen-binding region of the antibody (or its functional equivalent) and is most critical in specificity and affinity of binding. See Paul, *Fundamental Immunology* (2003).

[0065] An exemplary immunoglobulin (antibody) structural unit comprises a tetramer. Each tetramer is composed of two identical pairs of polypeptide chains, each pair having one "light" (about 25 kD) and one "heavy" chain (about 50-70 kD). The N-terminus of each chain defines a variable region of about 100 to 110 or more amino acids primarily responsible for antigen recognition. The terms variable light chain (V_L) and variable heavy chain (V_H) refer to these light and heavy chains respectively.

[0066] Antibodies can exist as intact immunoglobulins or as any of a number of well-characterized fragments that include specific antigen-binding activity. Such fragments can be produced by digestion with various peptidases. Pepsin digests an antibody below the disulfide linkages in the hinge region to produce $F(ab)'_2$, a dimer of Fab which itself is a light chain joined to V_H-C_H1 by a disulfide bond. The $F(ab)'_2$ may be reduced under mild conditions to break the disulfide linkage in the hinge region, thereby converting the $F(ab)'_2$ dimer into an Fab' monomer. The Fab' monomer is essentially Fab with part of the hinge region (*see Fundamental Immunology* (Paul ed., 3d ed. 1993). While various antibody fragments are defined in terms of the digestion of an intact antibody, one of skill will appreciate that such fragments may be synthesized *de novo* either chemically or by using recombinant DNA methodology. Thus, the term antibody, as used herein, also includes antibody fragments either produced by the modification of whole antibodies, or those synthesized *de novo* using recombinant DNA methodologies (*e.g.*, single chain Fv) or those identified using phage display libraries (*see, e.g.,* McCafferty *et al.*, *Nature* 348:552-554 (1990)).

II. Mucins

[0067] There are several gel-forming mucins including, but not limited to, MUC6, MUC2, MUC5AC, and MUC5B. These proteins are large filamentous and highly O-glycosylated.

III. Pulmonary Diseases

[0068] The pulmonary diseases contemplated herein can include any pulmonary disorders, lung fibrosis diseases, interstitial lung diseases, idiopathic interstitial pneumonias (IIP),

idiopathic pulmonary fibrosis, familial interstitial pneumonia (FIP), acute respiratory distress syndrome (ARDS), scleroderma lung disease , Sarcoidosis, Beryllium disease, rheumatoid arthritis associated lung disorder, collagen vascular associated lung disorder, cigarette smoke associated lung disorders, Sjögren's syndrome, mixed connective tissue disease, nonspecific interstitial pneumonitis (NSIP), etc.

[0069] Pulmonary fibrotic conditions, *e.g.*, interstitial lung diseases (ILD) are characterized by shortness of breath, chronic coughing, fatigue and weakness, loss of appetite, and rapid weight loss. Pulmonary fibrosis is commonly linked to interstitial lung diseases (*e.g.*, autoimmune disorders, viral infections or other microscopic injuries), but can be idiopathic. Fibrosis involves exchange of normal lung tissue with fibrotic tissue (scar tissue) that leads to reduced oxygen capacity.

[0070] Idiopathic interstitial pneumonias (IIP) are a subset of diffuse interstitial lung diseases of unknown etiology (the term "idiopathic" indicates unknown origin). IIPs are characterized by expansion of the interstitial compartment (*i.e.*, that portion of the lung parenchyma sandwiched between the epithelial and endothelial basement membranes) with an infiltrate of inflammatory cells. The inflammatory infiltrate is sometimes accompanied by fibrosis, either in the form of abnormal collagen deposition or proliferation of fibroblasts capable of collagen synthesis.

[0071] Idiopathic Pulmonary Fibrosis (IPF) occurs in thousands of people worldwide with a doubling of prevalence over the past 10 years. Onset of IPF occurs around 50 to 70 years of age and starts with progressive shortness of breath and hypoxemia. IPF median survival is around 3-5 years and is to date untreatable. The etiology and pathogenesis of the condition is not well understood. About 5-20 percent of all cases of IPF have a family history and inheritance appears to be autosomal dominant.

[0072] Additional fibrotic pulmonary diseases include Acute Interstitial Pneumonia (AIP), Respiratory Bronchiolitis-associated Interstitial Lung Disease (RBILD), Desquamative Interstitial Pneumonia (DIP), Non-Specific Interstitial Pneumonia (NSIP), Bronchiolitis obliterans, with Organizing Pneumonia (BOOP).

[0073] AIP is a rapidly progressive and histologically distinct form of interstitial pneumonia. The pathological pattern is an organizing form of diffuse alveolar damage (DAD) that is also found in acute respiratory distress syndrome (ARDS) and other acute interstitial pneumonias of known causes (*see* Clinical Atlas of Interstitial Lung Disease (2006 ed.) pp61-63).

[0074] RBILD is characterized by inflammatory lesions of the respiratory bronchioles in cigarette smokers. The histologic appearance of RBILD is characterized by the accumulation of pigmented macrophages within the respiratory bronchioles and the surrounding airspaces, variably, peribronchial fibrotic alveolar septal thickening, and minimal associated mural inflammation (*see* Wells *et al.* (2003) *Sem Respir. Crit. Care Med.* vol. 24).

[0075] DIP is a rare interstitial lung disease characterized by the accumulation of macrophages in large numbers in the alveolar spaces associated with interstitial inflammation and/or fibrosis. The macrophages frequently contain light brown pigment. Lymphoid nodules are common, as is a sparse but distinct eosinophil infiltrate. DIP is most common in smokers (*see* Tazelaar *et al.* (Sep. 21, 2010) *Histopathology*).

[0076] NSIP is characterized pathologically by uniform interstitial inflammation and fibrosis appearing over a short period of time. NSIP differs from other interstitial lung diseases in that it has a generally good prognosis. In addition, the temporal uniformity of the parenchymal changes seen in NSIP contrasts greatly with the temporal heterogeneity of usual interstitial pneumonia (*see* Coche *et al.* (2001) *Brit J Radiol* 74:189).

[0077] BOOP, unlike NSIP, can be fatal within days of first acute symptoms. It is characterized by rapid onset of acute respiratory distress syndrome; therefore, clinically, rapidly progressive BOOP can be indistinguishable from acute interstitial pneumonia. Histological features include clusters of mononuclear inflammatory cells that form granulation tissue and plug the distal airways and alveolar spaces. These plugs of granulation tissue may form polyps that migrate within the alveolar ducts or may be focally attached to the wall. (*see* White & Ruth-Saad (2007) *Crit. Care Nurse* 27:53).

[0078] Further details about the characteristics and therapies available for these diseases can be found, *e.g.*, on the website of the American Lung Association at lungusa.org/lung-disease/pulmonary-fibrosis.

[0079] Diagnostic indicators of pulmonary disorders include biopsy (*e.g.*, VATS or surgical lung biopsy), high resolution computed tomography (HRTC) or breathing metrics, such as forced expiratory volume (FEV1), vital capacity (VC), forced vital capacity (FVC), and FEV1/FVC.

[0080] Additional disorders associated with MUC5B expression and/or SNPs associated with *MUC5B* (*e.g.* SNP rs35705950) can include, but are not limited to, mucous secretion disorders, cancers (*e.g.* ovarian, breast lung, pancreatic etc.), eye disease, colitis, and cirrhosis of the liver.

IV. Methods of Diagnosis and Prognosis

[0081] Methods for detecting and identifying nucleic acids and proteins and interactions between such molecules involve conventional molecular biology, microbiology, and recombinant DNA techniques within the skill of the art. Such techniques are explained fully in the literature (*see, e.g.*, Sambrook, Fritsch & Maniatis, Molecular Cloning: A Laboratory Manual, Second Edition 1989, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.; Animal Cell Culture, R. I. Freshney, ed., 1986).

A. Biological samples

[0082] For detection of a genetic variant using genomic DNA, a biological sample can be obtained from nearly any tissue. One of skill in the art will understand that a blood sample or a cheek swab is expected to carry the same genetic sequence information as a lung cell. For detection of a given expression level, pulmonary tissue samples and other biological fluids are typically used.

[0083] Biological samples can include a pulmonary mucosal sample or biological fluid such as blood or blood components (plasma, serum), sputum, mucus, urine, saliva, etc.

[0084] A pulmonary mucosal sample can be obtained using methods known in the art, *e.g.*, a bronchial epithelial brush or exhaled breath condensate. Additional methods include bronchial biopsy, bronchial wash, bronchoalveolar lavage, whole lung lavage, transendoscopic biopsy, translaryngoscopic catheter, and transtracheal wash. A review of commonly used techniques, including comparisons and safety issues, is provided in Busse *et al.* (2005) *Am J Respir Crit Care Med* 172:807–816.

[0085] For lavage techniques, a bronchoscope can be inserted to the desired level of the airway. A small volume of sterile, physiologically acceptable fluid (*e.g.*, buffered saline) is released, and immediately aspirated. The wash material contains cells from the mucosa and upper epithelia (Riise *et al.* (1996) *Eur Resp J* 9:1665).

[0086] For use of a bronchial epithelial brush, a sterile, non-irritating (*e.g.*, nylon) cytology brush can be used. Multiple brushings can be taken to ensure representative sampling. The

brush is then agitated in physiologically acceptable fluid, and the cells and debris separated using routine methods (Riise *et al.* (1992) *Eur Resp J* 5:382).

[0087] Cellular components can be isolated using methods known in the art, *e.g.*, centrifugation. Similarly, subcellular components (*e.g.*, exosomes or vesicles) can be isolated using known methods or commercial separation products (available from BioCat, System Bio, Bioscientific, etc.). An exemplary method is described *e.g.*, by Thery *et al.* (2006) *Current Prot. Cell Biol.*

B. Detection of genetic variants

[0088] The inventors have found that genetic variations in the mucin genes are associated with pulmonary diseases. These genetic variations can be found in any part of the gene, *e.g.*, in the regulatory regions, introns, or exons. Relevant genetic variations may also be found the intergene regions, *e.g.*, in sequences between mucin genes. Insertions, substitutions, and deletions are included in genetic variants. Single nucleotide polymorphisms (SNPs) are exemplary genetic variants.

[0089] In particular, 14 independent SNPs are associated with pulmonary disorders (*e.g.* FIP or IPF). The studies disclosed herein demonstrate that presence of one or more of these SNPs associated with *MUC5B* can lead to predisposition to a pulmonary disorder. In addition, in some embodiments, if present, some of these SNPs are related to a transcription factor binding site. The transcription factor binding site can effect modulation of *MUC5B* expression, for example E2F3 loss, and HOXA9 and PAX-2 generation.

[0090] The invention thus provides methods for assessing the presence or absence of SNPs in a sample from a subject suspected of having or developing a pulmonary disorder (*e.g.*, because of family history). In certain embodiments, one or more SNPs are screened in one or more samples from a subject. The SNPs can be associated with one or more genes, *e.g.*, one or more *MUC* genes or other genes associated with mucous secretion. In some embodiments, a *MUC* gene associated SNP is associated with *MUC5B* and/or another *MUC* gene, such as *MUC5AC* or *MUC1*. SNPs contemplated for diagnostic, treatment, or prognosis can include SNPs found within a *MUC* gene and/or within a regulatory or promoter region associated with a *MUC* gene. For example, one or more SNPs can include, but are not limited to, detection of the SNPs of *MUC5B* shown in Table 4 (SEQ ID NOs:20-53), *e.g.*, SNP rs35705950 (SEQ ID NO:24), alone

or in combination with other genetic variations or SNPs and/or other diagnostic or prognostic methods.

[0091] Methods for detecting genetic variants such as a SNP are known in the art, *e.g.*, Southern or Northern blot, nucleotide array, amplification methods, etc. Primers or probes are designed to hybridize to a target sequence. For example, genomic DNA can be screened for the presence of an identified genetic element of using a probe based upon one or more sequences, *e.g.*, using a probe with substantial identity to a subsequence of the *MUC5B* gene, such as one of the subsequences shown in Table 4 (SEQ ID NOs: 20-53). Exemplary human *MUC5B* genomic sequences that can be used for reference and probe and primer design are found at GenBank Accession Nos. AF107890.1 and AJ004862.1. Expressed RNA can also be screened, but may not include all relevant genetic variations. Various degrees of stringency of hybridization may be employed in the assay. As the conditions for hybridization become more stringent, there must be a greater degree of complementarity between the probe and the target for duplex formation to occur. Thus, high stringency conditions are typically used for detecting a SNP.

[0092] Thus, in some embodiments, a genetic variant *MUC5B* gene in a subject is detected by contacting a nucleic acid in a sample from the subject with a probe having substantial identity to a subsequence of the *MUC5B* gene, and determining whether the nucleic acid indicates that the subject has a genetic variant *MUC5B* gene. In some cases, the sample can be processed prior to amplification, *e.g.*, to separate genomic DNA from other sample components. In some cases, the probe has at least 90, 92, 94, 95, 96, 98, 99, or 100% identity to the *MUC5B* gene subsequence. Typically, the probe is between 10-500 nucleotides in length, *e.g.*, 10-100, 10-40, 10-20, 20-100, 100-400, etc. In the case of detecting a SNP, the probe can be even shorter, e.g., 8-20 nucleotides in length. In some cases, the *MUC5B* gene sequence to be detected includes at least 8 contiguous nucleotides, *e.g.*, at least 10, 15, 20, 25, 30, 35 or more contiguous nucleotides of one of the sequences shown in SEQ ID NOs:20-53. In some embodiments, the sequence to be detected includes 8 contiguous nucleotides, *e.g.*, at least 10, 15, 20, 25, 30, 35 or more contiguous nucleotides of SEQ ID NO:24. In some aspects, the contiguous nucleotides include nucleotide 28 of SEQ ID NO:24.

[0093] The degree of stringency can be controlled by temperature, ionic strength, pH and/or the presence of a partially denaturing solvent such as formamide. For example, the stringency of hybridization is conveniently varied by changing the concentration of formamide within the

range up to and about 50%. The degree of complementarity (sequence identity) required for detectable binding will vary in accordance with the stringency of the hybridization medium and/or wash medium. In certain embodiments, in particular for detection of a particular SNP, the degree of complementarity is about 100 percent. In other embodiments, sequence variations can result in <100% complementarity, <90% complimentarity probes, <80% complimentarity probes, etc., in particular, in a sequence that does not involve a SNP. In some examples, *e.g.*, detection of species homologs, primers may be compensated for by reducing the stringency of the hybridization and/or wash medium.

[0094] High stringency conditions for nucleic acid hybridization are well known in the art. For example, conditions may comprise low salt and/or high temperature conditions, such as provided by about 0.02 M to about 0.15 M NaCl at temperatures of about 50°C to about 70°C. Other exemplary conditions are disclosed in the following Examples. It is understood that the temperature and ionic strength of a desired stringency are determined in part by the length of the particular nucleic acid(s), the length and nucleotide content of the target sequence(s), the charge composition of the nucleic acid(s), and by the presence or concentration of formamide, tetramethylammonium chloride or other solvent(s) in a hybridization mixture. Nucleic acids can be completely complementary to a target sequence or exhibit one or more mismatches.

[0095] Nucleic acids of interest (*e.g.*, nucleic acids comprising, or comprised within, SEQ ID NOs:20-53) can also be amplified using a variety of known amplification techniques. For instance, polymerase chain reaction (PCR) technology may be used to amplify target sequences (*e.g.*, genetic variants) directly from DNA, RNA, or cDNA. In some embodiments, a stretch of nucleic acids is amplified using primers on either side of a targeted genetic variation, and the amplification product is then sequenced to detect the targeted genetic variation (using, *e.g.*, Sanger sequencing, Pyrosequencing, Nextgen® sequencing technologies). For example, the primers can be designed to hybridize to either side of the upstream regulatory region of the *MUC5B* gene, and the intervening sequence determined to detect a SNP in the promoter region. In some embodiments, one of the primers can be designed to hybridize to the targeted genetic variant. In some cases, a genetic variant nucleotide can be identified using RT-PCR, *e.g.*, using labeled nucleotide monomers. In this way, the identity of the nucleotide at a given position can be detected as it is added to the polymerizing nucleic acid. The Scorpion™ system is a commercially available example of this technology.

[0096] Thus, in some embodiments, a genetic variant *MUC5B* gene in a subject is detected by amplifying a nucleic acid in a sample from the subject to form an amplification product, and determining whether the amplification product indicates a genetic variant *MUC5B* gene. In some cases, the sample can be processed prior to amplification, *e.g.*, to separate genomic DNA from other sample components. In some cases, amplifying comprises contacting the sample with amplification primers having substantial identity to *MUC5B* genomic subsequences, *e.g.*, at least 90, 92, 94, 95, 96, 98, 99, or 100% identity. Typically, the sequence to be amplified is between 30-1000 nucleotides in length, *e.g.*, 50-500, 50-400, 100-400, 50-200, 100-300, etc. In some cases, the sequence to be amplified or detected includes at least 8 contiguous nucleotides, *e.g.*, at least 10, 15, 20, 25, 30, 35 or more contiguous nucleotides of one of the sequences shown in SEQ ID NOs:20-53. In some embodiments, the sequence to be amplified or detected includes 8 contiguous nucleotides, *e.g.*, at least 10, 15, 20, 25, 30, 35 or more contiguous nucleotides of SEQ ID NO:24. In some aspects, the contiguous nucleotides include nucleotide 28 of SEQ ID NO:24.

[0097] Amplification techniques can also be useful for cloning nucleic acid sequences, to make nucleic acids to use as probes for detecting the presence of a target nucleic acid in samples, for nucleic acid sequencing, for control samples, or for other purposes. Probes and primers are also readily available from commercial sources, *e.g.*, from Invitrogen, Clonetech, etc.

C. Detection of expression levels

[0098] Expression of a given gene, *e.g.*, *MUC5B* or another mucin, pulmonary disease marker, or standard (control), is typically detected by detecting the amount of RNA (*e.g.*, mRNA) or protein. Sample levels can be compared to a control level.

[0099] Methods for detecting RNA are largely cumulative with the nucleic acid detection assays described above. RNA to be detected can include mRNA. In some embodiments, a reverse transcriptase reaction is carried out and the targeted sequence is then amplified using standard PCR. Quantitative PCR (qPCR) or real time PCR (RT-PCR) is useful for determining relative expression levels, when compared to a control. Quantitative PCR techniques and platforms are known in the art, and commercially available (*see, e.g.*, the qPCR Symposium website, available at qpcrsymposium.com). Nucleic acid arrays are also useful for detecting nucleic acid expression. Customizable arrays are available from, *e.g.*, Affimatrix. An exemplary

human MUC5B mRNA sequence, *e.g.*, for probe and primer design, can be found at GenBank Accession No. AF086604.1.

[0100] Protein levels can be detected using antibodies or antibody fragments specific for that protein, natural ligands, small molecules, aptamers, etc. An exemplary human MUC5B sequence, *e.g.*, for screening a targeting agent, can be found at UniProt Accession No. O00446.

[0101] Antibody based techniques are known in the art, and described, *e.g.*, in Harlow & Lane (1988) Antibodies: A Laboratory Manual and Harlow (1998) Using Antibodies: A Laboratory Manual; Wild, The Immunoassay Handbook, 3d edition (2005) and Law, Immunoassay: A Practical Guide (1996). The assay can be directed to detection of a molecular target (*e.g.*, protein or antigen), or a cell, tissue, biological sample, liquid sample or surface suspected of carrying an antibody or antibody target.

[0102] A non-exhaustive list of immunoassays includes: competitive and non-competitive formats, enzyme linked immunosorption assays (ELISA), microspot assays, Western blots, gel filtration and chromatography, immunochromatography, immunohistochemistry, flow cytometry or fluorescence activated cell sorting (FACS), microarrays, and more. Such techniques can also be used *in situ*, *ex vivo*, or *in vivo*, *e.g.*, for diagnostic imaging.

[0103] Aptamers are nucleic acids that are designed to bind to a wide variety of targets in a non-Watson Crick manner. An aptamer can thus be used to detect or otherwise target nearly any molecule of interest, including a pulmonary disease associated protein. Methods of constructing and determining the binding characteristics of aptamers are well known in the art. For example, such techniques are described in U.S. Patent Nos. 5,582,981, 5,595,877 and 5,637,459. Aptamers are typically at least 5 nucleotides, 10, 20, 30 or 40 nucleotides in length, and can be composed of modified nucleic acids to improve stability. Flanking sequences can be added for structural stability, *e.g.*, to form 3-dimensional structures in the aptamer.

[0104] Protein detection agents described herein can also be used as a treatment and/or diagnosis of pulmonary disease or predictor of disease progression, *e.g.*, propensity for survival, in a subject having or suspected of developing a pulmonary disorder. In certain embodiments, MUC5B antibodies can be used to assess MUC5B protein levels in a subject having or suspected of developing a pulmonary disorder. It is contemplated herein that antibodies or antibody fragments may be used to modulate MUC5B production in a subject having or suspected of developing a pulmonary disease. In certain embodiments, one or more agents capable of

modulating MUC5B may be used to treat a subject having or suspected of developing a pulmonary disorder. One or more antibodies or antibody fragments may be generated to detect one or more of the SNPs disclosed herein by any method known in the art.

[0105] In certain embodiments, MUC5B diagnostic tests may include, but are not limited to, alone or in combination, analysis of rs35705950 SNP in *MUC5B* gene, MUC5B mRNA levels, and/or MUC5B protein levels.

D. Additional pulmonary disease markers

[0106] The above methods of detection can be applied to additional pulmonary disease markers. That is, the expression level or presence of genetic variants of at least one additional pulmonary disease marker gene can be determined, or the activity of the marker protein can be determined, and compared to a standard control for the pulmonary disease marker. The examination of additional pulmonary disease markers can be used to confirm a diagnosis of pulmonary disease, monitor disease progression, or determine the efficacy of a course of treatment in a subject.

[0107] In some cases, pulmonary disease is indicated by an increased number of lymphocytes, *e.g.*, CD4+CD28- cells (Moeller *et al.* (2009) *Am. J. Resp. Crit Care. Med.* 179:588; Gilani (2010) *PLoS One* 5:e8959).

[0108] Genetic variations in the following genes are associated with pulmonary disease: Surfactant Protein A2, Surfactant Protein B, Surfactant Protein C, TERC, TERT, IL-1RN, IL-1α, IL-1β, TNF, Lymphotoxin α, TNF-RII, IL-10, IL-6, IL-12, IFNγ, TGFβ, CR1, ACE, IL-8, CXCR1, CXCR2, MUC1 (KL6), or MUC5AC. Thus, the invention further includes methods of determining whether the genome of a subject comprises a genetic variant of at least one gene selected from these genes. The presence of a genetic variant indicates that the subject has or is at risk of developing pulmonary disease. Said determining can optionally be combined with determining whether the genome of the subject comprises a genetic variant *MUC5B* gene, or determining whether the subject has an elevated level of MUC5B RNA or protein to confirm or strengthen the diagnosis or prognosis.

[0109] Abnormal expression in the following genes can also be indicative of pulmonary disease: Surfactant Protein A, Surfactant Protein D, KL-6/MUC1, CC16, CK-19, Ca 19-9, SLX, MCP-1, MIP-1a, ITAC, glutathione, type III procollagen peptide, sIL-2R, ACE, neopterin, beta-

glucuronidase, LDH, CCL-18, CCL-2, CXCL12, MMP7, and osteopontin. Thus, the expression of one of these genes can be detected and compared to a control, wherein an abnormal expression level indicates that the subject has or is at risk of developing pulmonary disease. Said determining can optionally be combined with determining whether the genome of the subject comprises a genetic variant *MUC5B* gene, or determining whether the subject has an elevated level of MUC5B RNA or protein to confirm or strengthen the diagnosis or prognosis.

E. Indications

[0110] The detection methods described herein can be used for diagnosis, prognosis, risk prediction, determining a course of treatment, monitoring therapeutic efficacy, and monitoring disease progression. One of skill will appreciate that each of the detection methods can be used alone or in combination.

[0111] For example, the presence of a genetic variant *MUC5B* gene can be determined in a subject suspected of having or at risk of developing a pulmonary disorder. In the event that a genetic variant *MUC5B* gene is observed, the subject can optionally undergo further testing, *e.g.*, to determine the level of *MUC5B* gene expression, or detect a genetic variant form of at least one additional pulmonary disease marker. The subject can be prescribed a course of treatment based on the results of one or more tests. Such treatment can include administration of a MUC5B antagonist, or a standard pulmonary disease treatment such as a mucolytic drug. The expression level of the *MUC5B* gene can be detected again after treatment, or periodically during the course of treatment, to determine the therapeutic efficacy of the treatment. For example, if a pulmonary disease treatment is prescribed for periodic administration (*e.g.*, daily, twice-daily, weekly, etc.), the *MUC5B* gene expression level can be monitored periodically thereafter (*e.g.*, monthly).

[0112] The detection methods of the invention can be used to determine if the subject has an attenuated form of the pulmonary disease. The inventors have shown that individuals carrying the rs35705950 genetic variant *MUC5B* gene have a better pulmonary disease prognosis than individuals that do not carry a genetic variant *MUC5B* gene. Thus, determination of whether an individual carries the genetic variant *MUC5B* gene can be used to design a course of treatment for the individual.

V. Methods of Treatment

A. Pulmonary disease treatments

[0113] A number of pulmonary disease treatments are available for addressing airway inflammation and/or excess mucus secretion. These include agents that can be roughly categorized, *e.g.*, as mucolytic agents, mucoregulatory agents, mucokinetic agents, and expectorants (*see, e.g.*, Balsamo *et al.* (2010) *Eur. Respir. Rev.* 19:127-33), though there is some overlap in the categories. Such agents are useful for treating the pulmonary diseases described herein, *e.g.*, as part of a course of treatment and monitoring, or after detection of elevated MUC5B RNA or protein, or detection of a genetic variant *MUC5B* gene.

[0114] Mucolytic drugs are those that decrease mucus viscosity, either by depolymerizing mucin glycoproteins or depolymerizing DNA and F-actin polymer networks. The first mode of action can be particularly useful for addressing excess MUC5B. Exemplary mucolytics include N-acetylcysteine, N-acystelyn, erdoseine, dornase alfa, thymosin beta4, dextran, pulmozyme, heparin, and bronchiotol (inhaled mannose).

[0115] Mucoregulators are those agents that regulate mucus secretion, or interfere with the DNA/ F-actin network. Examples of mucoregulators include, *e.g.*, carbocysteine, anticholoinergic agents, glucocorticoids, and macrolide antibiotics.

[0116] Mucokinetic agents increase mucus clearance by acting on the cilia lining the airway. Examplary mucokinetic agents include, *e.g.*, bronchodilators, surfactants, and ambroxol.

[0117] Expectorants are agents that induce discharge of mucus from the airway or respiratory tract. Some examples include hypertonic saline, guaifenesin, dornase/ pulmozyme, and bronchiotol (inhaled mannose).

[0118] The pulmonary disease treatment, such as the agents described above, can be used alone, sequentially, or in combination according to the methods described herein. In some embodiments, a pulmonary disease treatment is used in combination with a more targeted inhibitor of MUC5B expression.

B. MUC5B Antagonists

[0119] The results disclosed herein indicate that elevated expression of the *MUC5B* gene is associated with pulmonary disease. The invention thus includes methods and compositions for inhibiting the expression, secretion, and/ or activity of MUC5B. Exemplary inhibitors include

siRNA and antisense, pRNA (promoter-associated RNA, *see, e.g.*, Schmitz *et al.* (2010) *Genes Dev.* 24:2264-69), MUC5B-specific antibodies and fragments thereof, and MUC5B-specific aptamers. In some embodiments, MUC5B activity can be inhibited or MUC5B clearance can be increased, *e.g.*, using mucolytic agents, glycosylation inhibitors, or inhibitors of protein secretion. The terms "inhibitor" and "antagonist" and like terms are used synonymously herein.

[0120] Thus, a nucleotide sequence that specifically interferes with expression of the *MUC5B* gene at the transcriptional or translational level can be used to treat or prevent pulmonary disease. This approach may utilize, for example, siRNA and/or antisense oligonucleotides to block transcription or translation of a specific mRNA (*e.g.*, a genetic variant RNA), either by inducing degradation of the mRNA with a siRNA or by masking the mRNA with an antisense nucleic acid. In some embodiments, the siRNA or antisense construct does not significantly block expression of other mucin genes.

[0121] Double stranded siRNA that corresponds to the *MUC5B* gene can be used to silence the transcription and/or translation by inducing degradation of MUC5B mRNA transcripts, and thus treat or prevent pulmonary disease (e.g., pulmonary disease associated with genetic variant *MUC5B*). The siRNA is typically about 5 to about 100 nucleotides in length, more typically about 10 to about 50 nucleotides in length, most typically about 15 to about 30 nucleotides in length. siRNA molecules and methods of generating them are described in, *e.g.*, Bass, 2001, *Nature*, 411, 428-429; Elbashir *et al.*, 2001, *Nature*, 411, 494-498; WO 00/44895; WO 01/36646; WO 99/32619; WO 00/01846; WO 01/29058; WO 99/07409; and WO 00/44914. A DNA molecule that transcribes dsRNA or siRNA (for instance, as a hairpin duplex) also provides RNAi. DNA molecules for transcribing dsRNA are disclosed in U.S. Patent No. 6,573,099, and in U.S. Patent Application Publication Nos. 2002/0160393 and 2003/0027783, and Tuschl and Borkhardt, *Molecular Interventions*, 2:158 (2002). For example, dsRNA oligonucleotides that specifically hybridize to the *MUC5B* nucleic acid sequences described herein can be used in the methods of the present invention. A decrease in the severity of pulmonary disease symptoms in comparison to symptoms detected in the absence of the interfering RNA can be used to monitor the efficacy of the siRNA

[0122] Antisense oligonucleotides that specifically hybridize to nucleic acid sequences encoding MUC5B polypeptides can also be used to silence transcription and/or translation, and thus treat or prevent pulmonary disease. For example, antisense oligonucleotides that

specifically hybridize to a MUC5B polynucleotide sequence can be used. A decrease in the severity of pulmonary disease symptoms in comparison to symptoms detected in the absence of the antisense nucleic acids can be used to monitor the efficacy of the antisense nucleic acids.

[0123] Antisense nucleic acids are DNA or RNA molecules that are complementary to at least a portion of a specific mRNA molecule (*see*, *e.g.*, Weintraub, *Scientific American*, 262:40 (1990)). Typically, synthetic antisense oligonucleotides are generally between 15 and 25 bases in length. Antisense nucleic acids may comprise naturally occurring nucleotides or modified nucleotides such as, *e.g.*, phosphorothioate, methylphosphonate, and -anomeric sugar-phosphate, backbone-modified nucleotides.

[0124] In the cell, the antisense nucleic acids hybridize to the corresponding mRNA, forming a double-stranded molecule. The antisense nucleic acids, interfere with the translation of the mRNA, since the cell will not translate a mRNA that is double-stranded. Antisense oligomers of about 15 nucleotides are preferred, since they are easily synthesized and are less likely to cause problems than larger molecules when introduced into the target nucleotide mutant producing cell. The use of antisense methods to inhibit the in vitro translation of genes is well known in the art (Marcus-Sakura, *Anal. Biochem.*, 172:289, (1988)). Less commonly, antisense molecules which bind directly to the DNA may be used.

[0125] siRNA and antisense can be delivered to the subject using any means known in the art, including by injection, inhalation, or oral ingestion. Another suitable delivery system is a colloidal dispersion system such as, for example, macromolecule complexes, nanocapsules, microspheres, beads, and lipid-based systems including oil-in-water emulsions, micelles, mixed micelles, and liposomes. The preferred colloidal system of this invention is a liposome. Liposomes are artificial membrane vesicles which are useful as delivery vehicles in vitro and in vivo. Nucleic acids, including RNA and DNA within liposomes and be delivered to cells in a biologically active form (Fraley, *et al., Trends Biochem. Sci.*, 6:77, 1981). Liposomes can be targeted to specific cell types or tissues using any means known in the art.

[0126] The invention also provides antibodies that specifically bind to MUC5B protein. Such antibodies can be used to sequester secreted MUC5B, *e.g.*, to prevent gel-forming activity and formation of excess mucus.

[0127] An antibody that specifically detects MUC5B, and not other mucin proteins, can be isolated using standard techniques described herein. The protein sequences for MUC5B in a

number of species, *e.g.*, humans, non-human primates, rats, dogs, cats, horses, bovines, *etc.*, are publically available.

[0128] Monoclonal antibodies are obtained by various techniques familiar to those skilled in the art. Briefly, spleen cells from an animal immunized with a desired antigen are immortalized, commonly by fusion with a myeloma cell (*see*, for example, Kohler & Milstein, *Eur. J. Immunol.* 6: 511-519 (1976)). Alternative methods of immortalization include transformation with Epstein Barr Virus, oncogenes, or retroviruses, or other methods well known in the art. Colonies arising from single immortalized cells are screened for production of antibodies of the desired specificity and affinity for the antigen, and yield of the monoclonal antibodies produced by such cells may be enhanced by various techniques, including injection into the peritoneal cavity of a vertebrate host. Alternatively, one may isolate DNA sequences which encode a monoclonal antibody or a binding fragment thereof by screening a DNA library from human B cells according to the general protocol outlined by Huse *et al.*, *Science* 246: 1275-1281 (1989).

[0129] Monoclonal antibodies are collected and titered against the MUC5B in an immunoassay, for example, a solid phase immunoassay with the immunogen immobilized on a solid support. Monoclonal antibodies will usually bind with a K_d of at least about 0.1 mM, more usually at least about 1 μM, and can often be designed to bind with a K_d of 1nM or less.

[0130] The immunoglobulins, including MUC5B-binding fragments and derivatives thereof, can be produced readily by a variety of recombinant DNA techniques, including by expression in transfected cells (*e.g.*, immortalized eukaryotic cells, such as myeloma or hybridoma cells) or in mice, rats, rabbits, or other vertebrate capable of producing antibodies by well known methods. Suitable source cells for the DNA sequences and host cells for immunoglobulin expression and secretion can be obtained from a number of sources, such as the American Type Culture Collection (Catalogue of Cell Lines and Hybridomas, Fifth edition (1985) Rockville, Md).

[0131] In some embodiments, the antibody is a humanized antibody, *i.e.*, an antibody that retains the reactivity of a non-human antibody while being less immunogenic in humans. This can be achieved, for instance, by retaining the non-human CDR regions that are specific for MUC5B, and replacing the remaining parts of the antibody with their human counterparts. *See, e.g.*, Morrison *et al.*, *PNAS USA*, 81:6851-6855 (1984); Morrison and Oi, *Adv. Immunol.*, 44:65-92 (1988); Verhoeyen *et al.*, *Science*, 239:1534-1536 (1988); Padlan, *Molec. Immun.*, 28:489-498 (1991); Padlan, *Molec. Immun.*, 31(3):169-217 (1994). Techniques for humanizing

antibodies are well known in the art and are described in *e.g.*, U.S. Patent Nos. 4,816,567; 5,530,101; 5,859,205; 5,585,089; 5,693,761; 5,693,762; 5,777,085; 6,180,370; 6,210,671; and 6,329,511; WO 87/02671; EP Patent Application 0173494; Jones *et al.* (1986) *Nature* 321:522; and Verhoyen *et al.* (1988) *Science* 239:1534. Humanized antibodies are further described in, *e.g.*, Winter and Milstein (1991) *Nature* 349:293. For example, polynucleotides comprising a first sequence coding for humanized immunoglobulin framework regions and a second sequence set coding for the desired immunoglobulin complementarity determining regions can be produced synthetically or by combining appropriate cDNA and genomic DNA segments. Human constant region DNA sequences can be isolated in accordance with well known procedures from a variety of human cells.

[0132] The activity of MUC5B protein can be inhibited, or the clearance of MUC5B can be increased, using mucolytic agents that break up mucus and proteolyze mucins. Mucolytic agents are described herein. Additional inhibitors of MUC5B protein include glycosylation inhibitors and inhibitors of protein secretion from epithelial cells. An exemplary glycosylation inhibitor includes benzyl-O-N-acetyl-D galactosamine (specific for O-glycans) and . Additional inhibitors of protein glycosylation are disclosed, *e.g.*, in Jacob (1995) *Curr. Opin. Structural Biol.* 5:605-11 and Patsos *et al.* 2005 *Biochem Soc. Trans.* 33:721-23. Secretion inhibitors include Brefeldin A, colchicine, and small molecules such as that disclosed in Stockwell (2006) *Nat. Chem. Biol.* 2:7-8. MUC5B activity can also be modulated by targeting the MARCKS protein (Adler *et al.* (2000) *Chest* 117: Supp 1 266S-267S).

C. Methods of Identifying MUC5B Antagonists

[0133] The invention further provides methods for identifying additional antagonists of MUC5B expression, secretion, and/or activity. Methods for screening for antagonists can involve measuring the ability of the potential antagonists to reduce an identifiable MUC5B activity or compete for binding with a known binding agent (*e.g.*, MUC5B-specific antibody). For example, candidate agents can be screened for their ability to reduce MUC5B gel formation, reduce MUC5B secretion, reduce MUC5B glycosylation, etc.

[0134] The screening methods of the invention can be performed as *in vitro* or cell-based assays. Cell based assays can be performed in any cells in which MUC5B is expressed, either endogenously or through recombinant methods. Cell-based assays may involve whole cells or cell fractions containing MUC5B to screen for agent binding or modulation of MUC5B activity

by the agent. Suitable cell-based assays are described in, *e.g.*, DePaola *et al.*, *Annals of Biomedical Engineering* 29: 1-9 (2001).

[0135] Agents that are initially identified as inhibiting MUC5B can be further tested to validate the apparent activity. Preferably such studies are conducted with suitable cell-based or animal models of pulmonary disease. The basic format of such methods involves administering a lead compound identified during an initial screen to an animal that serves as a model and then determining if in fact the pulmonary disease is ameliorated. The animal models utilized in validation studies generally are mammals of any kind. Specific examples of suitable animals include, but are not limited to, primates (*e.g.*, chimpanzees, monkeys, and the like) and rodents (*e.g.*, mice, rats, guinea pigs, rabbits, and the like).

[0136] The agents tested as potential antagonists of MUC5B can be any small chemical compound, or a biological entity, such as a polypeptide, sugar, nucleic acid or lipid. Alternatively, modulators can be genetically altered versions of MUC5B, *e.g.*, forms that are not glycosylated. Essentially any chemical compound can be used as a potential modulator or ligand in the assays of the invention, although most often compounds that can be dissolved in aqueous or organic (especially DMSO-based) solutions are used. The assays are designed to screen large chemical libraries by automating the assay steps and providing compounds from any convenient source to assays, which are typically run in parallel (*e.g.*, in microtiter formats on microtiter plates in robotic assays).

[0137] In one embodiment, high throughput screening methods involve providing a combinatorial chemical or peptide library containing a large number of potential therapeutic compounds (potential modulator or ligand compounds). Such "combinatorial chemical libraries" or "ligand libraries" are then screened in one or more assays, as described herein, to identify those library members (particular chemical species or subclasses) that display a desired characteristic activity. The compounds thus identified can serve as conventional "lead compounds" or can themselves be used as potential or actual therapeutics.

[0138] A combinatorial chemical library is a collection of diverse chemical compounds generated by either chemical synthesis or biological synthesis, by combining a number of chemical "building blocks" such as reagents. For example, a linear combinatorial chemical library such as a polypeptide library is formed by combining a set of chemical building blocks (amino acids) in every possible way for a given compound length (*i.e.*, the number of amino

acids in a polypeptide compound). Millions of chemical compounds can be synthesized through such combinatorial mixing of chemical building blocks.

[0139] Preparation and screening of combinatorial chemical libraries is well known to those of skill in the art. Such combinatorial chemical libraries include, but are not limited to, peptide libraries (*see, e.g.,* U.S. Patent 5,010,175, Furka, *Int. J. Pept. Prot. Res.* 37:487-493 (1991) and Houghton *et al.*, *Nature* 354:84-88 (1991)). Other chemistries for generating chemical diversity libraries can also be used. Such chemistries include, but are not limited to: peptoids (*e.g.*, PCT Publication No. WO 91/19735), encoded peptides (*e.g.*, PCT Publication WO 93/20242), random bio-oligomers (*e.g.*, PCT Publication No. WO 92/00091), benzodiazepines (*e.g.*, U.S. Pat. No. 5,288,514), diversomers such as hydantoins, benzodiazepines and dipeptides (Hobbs *et al.*, *Proc. Nat. Acad. Sci. USA* 90:6909-6913 (1993)), vinylogous polypeptides (Hagihara *et al.*, *J. Amer. Chem. Soc.* 114:6568 (1992)), nonpeptidal peptidomimetics with glucose scaffolding (Hirschmann *et al.*, *J. Amer. Chem. Soc.* 114:9217-9218 (1992)), analogous organic syntheses of small compound libraries (Chen *et al.*, *J. Amer. Chem. Soc.* 116:2661 (1994)), oligocarbamates (Cho *et al.*, *Science* 261:1303 (1993)), and/or peptidyl phosphonates (Campbell *et al.*, *J. Org. Chem.* 59:658 (1994)), nucleic acid libraries (*see* Ausubel, Berger and Sambrook, all *supra*), peptide nucleic acid libraries (*see, e.g.,* U.S. Patent 5,539,083), antibody libraries (*see, e.g.,* Vaughn *et al.*, *Nature Biotechnology*, 14(3):309-314 (1996) and PCT/US96/10287), carbohydrate libraries (*see, e.g.,* Liang *et al.*, *Science*, 274:1520-1522 (1996) and U.S. Patent 5,593,853), small organic molecule libraries (*see, e.g.,* benzodiazepines, Baum C&EN, Jan 18, page 33 (1993); isoprenoids, U.S. Patent 5,569,588; thiazolidinones and metathiazanones, U.S. Patent 5,549,974; pyrrolidines, U.S. Patents 5,525,735 and 5,519,134; morpholino compounds, U.S. Patent 5,506,337; benzodiazepines, and US Patent No. 5,288,514).

D. Pharmaceutical compositions

[0140] The compositions disclosed herein can be administered by any means known in the art. For example, compositions may include administration to a subject intravenously, intradermally, intraarterially, intraperitoneally, intralesionally, intracranially, intraarticularly, intraprostaticaly, intrapleurally, intratracheally, intranasally, intravitreally, intravaginally, intrarectally, topically, intratumorally, intramuscularly, intrathecally, subcutaneously, subconjunctival, intravesicularlly, mucosally, intrapericardially, intraumbilically, intraocularly, orally, locally, by inhalation, by injection, by infusion, by continuous infusion, by localized perfusion, via a catheter, via a lavage,

in a creme, or in a lipid composition. Administration can be local, *e.g.*, to the pulmonary mucosa, or systemic.

[0141] Solutions of the active compounds as free base or pharmacologically acceptable salt can be prepared in water suitably mixed with a surfactant, such as hydroxypropylcellulose. Dispersions can also be prepared in glycerol, liquid polyethylene glycols, and mixtures thereof and in oils. Under ordinary conditions of storage and use, these preparations can contain a preservative to prevent the growth of microorganisms.

[0142] Pharmaceutical compositions can be delivered via intranasal or inhalable solutions or sprays, aerosols or inhalants. Nasal solutions can be aqueous solutions designed to be administered to the nasal passages in drops or sprays. Nasal solutions can be prepared so that they are similar in many respects to nasal secretions. Thus, the aqueous nasal solutions usually are isotonic and slightly buffered to maintain a pH of 5.5 to 6.5. In addition, antimicrobial preservatives, similar to those used in ophthalmic preparations, and appropriate drug stabilizers, if required, may be included in the formulation. Various commercial nasal preparations are known and can include, for example, antibiotics and antihistamines.

[0143] Oral formulations can include excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate and the like. These compositions take the form of solutions, suspensions, tablets, pills, capsules, sustained release formulations or powders. In some embodiments, oral pharmaceutical compositions will comprise an inert diluent or assimilable edible carrier, or they may be enclosed in hard or soft shell gelatin capsule, or they may be compressed into tablets, or they may be incorporated directly with the food of the diet. For oral therapeutic administration, the active compounds may be incorporated with excipients and used in the form of ingestible tablets, buccal tablets, troches, capsules, elixirs, suspensions, syrups, wafers, and the like. Such compositions and preparations should contain at least 0.1% of active compound. The percentage of the compositions and preparations may, of course, be varied and may conveniently be between about 2 to about 75% of the weight of the unit, or preferably between 25-60%. The amount of active compounds in such compositions is such that a suitable dosage can be obtained

[0144] For parenteral administration in an aqueous solution, for example, the solution should be suitably buffered and the liquid diluent first rendered isotonic with sufficient saline or glucose. Aqueous solutions, in particular, sterile aqueous media, are especially suitable for

intravenous, intramuscular, subcutaneous and intraperitoneal administration. For example, one dosage could be dissolved in 1 ml of isotonic NaCl solution and either added to 1000 ml of hypodermoclysis fluid or injected at the proposed site of infusion

[0145] Sterile injectable solutions can be prepared by incorporating the active compounds or constructs in the required amount in the appropriate solvent followed by filtered sterilization. Generally, dispersions are prepared by incorporating the various sterilized active ingredients into a sterile vehicle which contains the basic dispersion medium. Vacuum-drying and freeze-drying techniques, which yield a powder of the active ingredient plus any additional desired ingredients, can be used to prepare sterile powders for reconstitution of sterile injectable solutions. The preparation of more, or highly, concentrated solutions for direct injection is also contemplated. DMSO can be used as solvent for extremely rapid penetration, delivering high concentrations of the active agents to a small area.

E. Treatment regimes

[0146] The invention provides methods of treating, preventing, and/or ameliorating a pulmonary disorder in a subject in need thereof, optionally based on the diagnostic and predictive methods described herein. The course of treatment is best determined on an individual basis depending on the particular characteristics of the subject and the type of treatment selected. The treatment, such as those disclosed herein, can be administered to the subject on a daily, twice daily, bi-weekly, monthly or any applicable basis that is therapeutically effective. The treatment can be administered alone or in combination with any other treatment disclosed herein or known in the art. The additional treatment can be administered simultaneously with the first treatment, at a different time, or on an entirely different therapeutic schedule (*e.g.*, the first treatment can be daily, while the additional treatment is weekly).

[0147] Administration of a composition for ameliorating the pulmonary disease, *e.g.*, by treating elevated expression of the *MUC5B* gene, can be a systemic or localized administration. For example, treating a subject having a pulmonary disorder can include administering an inhalable or intranasal form of anti-MUC5B agent (MUC5B antagonist) on a daily basis or otherwise regular schedule. In some embodiments, the treatment is only on an as-needed basis, *e.g.*, upon appearance of pulmonary disease symptoms.

VI. Kits

[0148] The invention provides kits for detection of pulmonary disease markers in a subject. The kit can be for personal use or provided to medical professionals. The kit can be a kit for diagnosing or prognosing a pulmonary disorder, or for monitoring the progression of disease or the efficacy of treatment.

[0149] In some embodiments, the kit includes components for assessing *MUC5B* gene expression comprising, *e.g.*, a nucleic acid capable of detecting MUC5B RNA or a MUC5B protein binding agent, optionally labeled. One of skill will appreciate that MUC5B gene expression can be determined by measuring MUC5B RNA or protein. The kit can further include assay containers (tubes), buffers, or enzymes necessary for carrying out the detection assay.

[0150] In some embodiments, the kit includes components for determining whether the genome of the subject carries a genetic variant *MUC5B* gene, *e.g.*, a nucleic acid that specifically hybridizes to a genetic variant *MUC5B* gene sequence. Other components in a kit can include, DNA sequencing assay components, Taqman ® genotyping assay components, Meta Analysis, one or more detection system(s), one or more control samples or a combination thereof. Kits can further include one or more agents where at least one of the agents is capable of associating with SNP rs35705950.

[0151] In some embodiments, the kit includes components to examine more than one pulmonary disease marker. For example, the kit can include marker detection agents, such as marker specific primers or probes attached to an addressable array. Exemplary markers include SNPs in the *MUC5B* genes, or genetic variants in other genes, e.g., Surfactant Protein A2, Surfactant Protein B, Surfactant Protein C, TERC, TERT, IL-1RN, IL-1α, IL-1β, TNF, Lymphotoxin α, TNF-RII, IL-10, IL-6, IL-12, IFNγ, TGFβ, CR1, ACE, IL-8, CXCR1 or CXCR2. In some embodiments, the expression level of the markers is detected instead of or in addition to the genetic sequence. In this case, useful pulmonary disease markers with aberrant expression include: Surfactant Protein A, Surfactant Protein D, KL-6/MUC1, CC16, CK-19, Ca 19-9, SLX, MCP-1, MIP-1a, ITAC, glutathione, type III procollagen peptide, sIL-2R, ACE, neopterin, beta-glucuronidase, LDH, CCL-18, CCL-2, CXCL12, MMP7, and osteopontin. Additional pulmonary disease markers can include the other MUC genes, *e.g.*, MUC2, MUC5AC, and MUC6.

[0152] The kit will generally include at least one vial, test tube, flask, bottle, syringe or other container means, into which the testing agent, can be suitably reacted or aliquoted. Kits can also include components for comparing results such as a suitable control sample, for example a positive and/or negative control. The kit can also include a collection device for collecting and/or holding the sample from the subject. The collection device can include a sterile swab or needle (for collecting blood), and/or a sterile tube (*e.g.*, for holding the swab or a bodily fluid sample).

[0153] The following discussion of the invention is for the purposes of illustration and description, and is not intended to limit the invention to the form or forms disclosed herein. Although the description of the invention has included description of one or more embodiments and certain variations and modifications, other variations and modifications are within the scope of the invention, *e.g.*, as may be within the skill and knowledge of those in the art, after understanding the present disclosure. All publications, patents, patent applications, Genbank numbers, and websites cited herein are hereby incorporated by reference in their entireties for all purposes.

VII. Examples

Example 1: Sequencing of pulmonary, gel-forming mucins and disease association

[0154] *Study populations:* Subjects with FIP or IPF were identified and phenotyped. The diagnosis of IIP was established according to conventional criteria. Eligible subjects were at least 38 years of age and had IIP symptoms for at least 3 months. A high resolution computerized tomography (HRCT) scan was required to show definite or probable IIP according to predefined criteria, and a surgical lung biopsy was obtained in 46% of affected subjects. FIP families were defined by the presence of two or more cases of definite or probable IIP within three degrees, with at least one case of IIP established as definite/probable IPF. Exclusion criteria included significant exposure to known fibrogenic agents or an alternative etiology for ILD. Control subjects for genetic analysis were acquired (Fig. 1).

[0155] *Linkage analysis:* A genome-wide linkage screen was completed in 82 multiplex families using a DeCode linkage panel consisting of a total of 884 markers with an average inter-marker distance of 4.2 CM. Multipoint non-parametric linkage analysis was performed using Merlin, previously described. Kong and Cox LOD scores were calculated using the S_{pairs} statistic

under an exponential model; support intervals were determined using the one-LOD-score-down method.

[0156] *Fine-mapping of chromosome 11:* To interrogate the linked region on the p-terminus of chromosome 11 (8.4 Mb bounded by rs702966 and rs1136966), fine mapping by genotyping 306 tagging SNPs in 145 unrelated cases of FIP, 152 cases of IPF, and 233 Caucasian controls were performed. Tests of association comparing FIP cases and IPF cases to controls were calculated under an additive model for the minor allele.

[0157] *Resequencing of MUC2 and MUC5AC:* Primer pairs to generate overlapping amplicons for resequencing the proximal promoter and most exons of *MUC2* and *MUC5AC* were designed on sequences masked for repetitive elements, SNPs, and homology to other regions of the genome.

[0158] *Genetic screen of lung-expressed, gel-forming mucins:* A case-control association study was conducted in an independent population of FIP (N=83), sporadic IPF (N=492), and control (N=322) subjects (Table 2) using tagging and other SNPs localized across the lung-expressed, gel-forming mucin genes on chromosome 11. 175 SNPs were successfully genotyped using the Sequenom iPlex assays, and haploview was used to test SNPs for allelic association with FIP and IPF. For those SNPs remaining significant after Bonferroni correction, odds ratios were estimated under an additive model for the rare allele after adjustment for age and gender via logistic regression. Chi-squared goodness-of-fit tests were computed to evaluate the evidence for disease-model explanations for genotypic departures from Hardy Weinberg Equilibrium (HWE) among cases. For the most highly-associated SNP, linkage and association modeling in pedigrees were used to test whether, in the original linkage families, the SNP was linked to the disease locus, was in linkage disequilibrium with the disease locus, and could account for the linkage signal.

[0159] Strong evidence for linkage based on the 82 FIP families occurred on chromosome 11 where the maximum multipoint LOD score was 3.3 (p=0.00004, D11S1318; Fig. 3). The 1-LOD support interval for this linked region was bounded by markers D11S4046 and D11S1760, spanning 3.4 Mb. Since D11S4046 was the most telomeric marker typed, the region of interest was inclusive of the p-terminus of chromosome 11. Within the 8.4 Mb larger region, 306 tagging SNPs were selected for fine-mapping in a case-control association analysis (145 FIP cases, 152 IPF cases, and 233 controls. Allelic association testing revealed 7 SNPs within the

mucin 2 (*MUC2*) gene significantly associated with either FIP or IPF. *MUC2* is contained in a genomic region harboring 4 gel-forming mucin genes (telomere to centromere: *MUC6, MUC2, MUC5AC*, and *MUC5B*). While there are reported recombination hotspots located between *MUC6* and *MUC2*, and within the proximal portion of *MUC5B*, markers within *MUC2* and *MUC5AC* exhibit strong linkage disequilibrium (LD) 17. Thus, *MUC2* and *MUC5AC* were selected for resequencing using the oligonucleotide primers. Resequencing analysis identified 330 genetic variants in *MUC2* and 195 genetic variants in *MUC5AC*. Allelic association testing between these genetic variants and disease status yielded 7 independent SNPs in both *MUC2* and *MUC5AC* significantly associated with either FIP or IPF disease status.

[0160] We designed a genetic screen for common genetic variation across the genomic region containing the 3 gel-forming mucin genes expressed in the lung (*MUC2, MUC5AC*, and *MUC5B*) in an independent population of subjects with IIP (FIP=83 and IPF=492) and controls (n=322) (Fig. 1, Table 2). 19 independent SNPs were observed to be significantly associated by allelic test with either or both FIP or IPF after Bonferroni correction for multiple comparisons (Table 1). Of these 19 SNPs, 6 occurred in *MUC2*, one in the *MUC2-MUC5AC* intergenic region, 4 in *MUC5AC*, 3 in the *MUC5AC-MUC5B* intergenic region, and 5 in the putative *MUC5B* promoter, within 4kb of the *MUC5B* transcription start site 18, 19 (Table 1).

[0161] Of significance, a SNP in the putative promoter of *MUC5B*, 3kb upstream of the transcription start site (rs35705950) was found to have the most substantial effect on both FIP and IPF. The minor allele of this SNP was present at a frequency of 33.8% in FIP cases, 37.5% in IPF cases, and 9.1% among controls (allelic association; FIP P=1.2x10-15, IPF P=2.5x10-37). Notably, the genotype frequencies for rs35705950 were consistent with HWE in controls, but not among IPF cases (P=6.0x10-11) and nearly so among FIP cases (P=0.11). By comparing the genotype frequencies observed in cases and controls to those expected if rs35705950 is a true risk locus, the data demonstrates that these genotype frequencies are consistent with an additive genotypic effect on disease risk conferred by rs35705950 (P=0.88 and P=0.77, respectively for FIP and IPF to reject additive effect). In addition, the disease allele frequency and penetrance estimates suggest a similar disease model for both FIP and IPF. The odds ratio for disease for subjects heterozygous and homozygous for the rarer allele of this SNP were 6.8 (95% CI 3.9-12.0) and 20.8 (95% CI 3.8-113.7) for FIP, and 9.0 (95% CI 6.2-13.1) and 21.8 (95% CI 5.1-93.5) for IPF (Table 1). To ensure this SNP was not tagging another SNP in the *MUC5B* promoter region, the 4kb region was resequenced upstream of the *MUC5B* transcription start site

in 48 IPF cases and 48 controls (Table 3). It was observed that 34 genetic variants but none had a pairwise r2 LD value with rs35705950 above 0.2 (Table 4). Finally, among the original linkage families, rs35705950 was found to be both linked to (P=0.04) and in linkage disequilibrium with (P=1.5x10-9) the disease locus. While there is some evidence for other linked variants in the region (P=0.054), these results verify the relevance of this SNP to disease in these families.

Table 1. Genotypic association results assuming an additive model from the genetic screen of lung-expressed, gel-forming mucins in subjects with IIP (FIP=83 and IPF=492) and controls (n=322).

SNP	Nucleotide Amino Acid Change	Mucin Region	Hg19 Position	Minor Allele Frequency			Genotypic Association Test by Disease Group			
				FIP (n=83)	IPF (n=492)	Controls (n=322)	Odds Ratio (95%CI)	PValue	Odds Ratio (95%CI)	PValue
rs10902081	C/T	MUC2 Int7	1079809	37.2	38.6	47.9	0.6(0.4-0.9)	0.011	0.7(0.5-0.8)	$4.3x10^{-4}$
rs7127117*	T/C	MUC2 Int7	1079879	49.3	60	47.4	1.0(0.7-1.5)	0.826	1.6(1.3-2.0)	$6.9x10^{-5}$
rs4145346	C/T Tyr426Tyr	MUC2 Ex10	1080894	5	6.5	2.2	1.9(0.8-4.3)	0.124	2.8(1.6-5.2)	0.001
rs4148348	G/A Thr618Thr	MUC2 Ex15	1082605	8.4	6.5	12.1	0.7(0.4-1.2)	0.188	0.5(0.4-0.8)	0.001
rs7934606*	C/T	MUC2 Int31	1098945	49.4	54	40.5	1.4(1.0-2.0)	0.055	1.7(1.4-2.2)	$3.8x10^{-6}$
rs10902089*	A/G	MUC2 Int31	1099657	57.9	58.8	48.5	1.5(1.0-2.1)	0.031	1.5(1.2-1.9)	$2.9x10^{-4}$
rs866739	C/T	MUC2-5AC Intergenic	1143101	22.5	21	12.5	2.2(1.4-3.6)	0.001	1.9(1.4-2.7)	$5.6x10^{-5}$
rs5584609	G/A Arg47Gln	MUC5AC Ex2	1154294	3.1	5.5	1.6	1.7(0.6-5.1)	0.316	3.6(1.7-7.3)	0.001
rs2240557	C/T Ala497Val	MUC5AC Ex12	1161315	8.9	13	3.4	2.7(1.3-5.3)	0.006	4.6(2.8-7.6)	$3.2x10^{-9}$
MUC5AC02547*	C/T	MUC5AC Int26	826476**	20.1	21	13.8	1.6(1.0-2.5)	0.053	1.6(1.2-2.2)	0.003
rs3528861	G/T	MUC5AC Int46	1220462	28.8	26.6	15.9	2.2(1.4-3.5)	$3.2x10^{-4}$	2.0(1.5-2.6)	$3.7x10^{-6}$
rs3571223	C/T	MUC5AC-5B Intergenic	1227089	42.6	42.4	33.4	1.4(1.0-2.0)	0.05	1.5(1.2-1.9)	0.001
rs2864232	C/T	MUC5AC-5B Intergenic	1229227	21.6	22.8	32.9	0.6(0.4-0.9)	0.009	0.6(0.5-0.8)	$1.1x10^{-4}$
rs3466903*	C/T	MUC5AC-5B Intergenic	1230883	21.5	23.3	34.8	0.5(0.3-0.7)	0.001	0.5(0.4-0.7)	$2.4x10^{-6}$
rs2672794	C/T	MUC5B Prm	1241005	27.2	27.5	40.4	0.5(0.3-0.8)	0.001	0.5(0.4-0.7)	$1.9x10^{-7}$
rs3570550	G/T	MUC5B Prm	1241221	33.8	37.5	9.1	6.2(3.7-10.4)	$3.7x10^{-12}$	8.3(5.8-11.9)	$4.6x10^{-34}$
rs3661943*	G/T	MUC5B Prm	1242250	40.3	39	23.8	2.4(1.6-3.6)	$3.3x10^{-5}$	2.1(1.6-2.8)	$1.5x10^{-8}$
rs1280404	G/T	MUC5B Prm	1242299	39.2	39.4	48.9	0.6(0.4-0.9)	0.019	0.6(0.5-0.8)	$1.2x10^{-4}$
rs868903*	T/C	MUC5B Prm	1242690	65.4	61	49.5	1.8(1.3-2.6)	0.001	1.6(1.3-2.1)	$2.8x10^{-5}$

* For these SNPs, DNA was available for 304 controls.** Nucleotide position based on NW_001838016.1.

Table 2. Demographic characteristics of subjects in the re-sequencing and mucin genetic screen analyses.

	Re-Sequencing Subjects			Genetic Screen of Lung-expressed Gel-forming Mucins Subjects		
	FIP	**IPF**	**Control**	**FIP**	**IPF**	**Control**
Number of subjects	69	96	54	83	492	322*
Male gender	41 (60%)	61 (64%)	18 (34%)	44 (53.0%)	352 (71.5%)	147 (45.7%)
Caucasian	68 (99%)	89 (93%)	53 (98%)	83 (100%)	492 (100%)	322 (100%)

Age at diagnosis	66 ± 10	65 ± 8	68 ± 8	66.3± 11.2	67.2 ± 8.1	60.3 ± 12.6
Ever smoked	44 (64%)	71 (74%)	25 (47%)	46 (56.8%)	342 (69.9%)	245 (76.6%)

* 325 control subjects were included in allelic association analyses but only 322 in genotypic regression analyses as demographic variables needed for regression were missing for 3 subjects. Additionally, in some genotyping multiplexes for the lung-expressed gel forming mucins, 18 of the 322 controls were not screened due to lack of DNA availability

Table 3: Oligos used in resequencing of the *MUC5B* promoter

MUC5B Promoter Amplicon	Forward Primer 5' > 3'	Reverse Primer 5' > 3'	Amplicon Size (bp)	Hg19 Coordinates
MUC5B-Prim-1	GGTTCTCCTTGTCTTGCAGCCCCT (SEQ ID NO:1)	ATGGGCTCTTGGTCTGCTCAGAG (SEQ ID NO:2)	616	Chr11:1239997-1240612
MUC5B-Prim-2	GGGCCTGGCTCTGAGTACACATCCT (SEQ ID NO:3)	AAGGAAAGGGACACAGCCGGTTCC (SEQ ID NO:4)	644	Chr11:1240556-1241199
MUC5B-Prim-3	GGGTCCCCATTCATGGCAGGATT (SEQ ID NO:5)	TTTCTCCATGGCAGAGCTGGGACC (SEQ ID NO:6)	601	Chr11:1240957-1241557
MUC5B-Prim-4	CTAGTGGGAGGGACGAGGGCAAAGT (SEQ ID NO:7)	CTCGTGGCTGTGACTGCACCCAG (SEQ ID NO:8)	610	Chr11:1241386-1241995
MUC5B-Prim-5	TTGGCTAAGGTGGGAGACCT (SEQ ID NO:9)	AGCTTGGGAATGTGAGAACG (SEQ ID NO:10)	700	Chr11:1241791-1242490
MUC5B-Prim-6	CATGAGGGGTGACAGGTGGCAAA (SEQ ID NO:11)	CCCGCGTTTGTCTTTCTGAAGTT (SEQ ID NO:12)	676	Chr11:1242392-1243067
MUC5B-Prim-7	GGTCAGAAGCTTTGAAGATGGGC (SEQ ID NO:13)	CTTGTCCAATGCCAGCCCTGATC (SEQ ID NO:14)	607	Chr11:1242985-1243591
MUC5B-Prim-8	CTGCCAGGGTTAATGAGGAG (SEQ ID NO:14)	GGATCAGGAAGGATTTGCAG (SEQ ID NO:16)	663	Chr11:1243491-1244153
MUC5B-Prim-9	AGGCAGGCTGGCTGACCACTGTTT (SEQ ID NO:17)	CGTGAAGACAGCATCGAGAGGGG (SEQ ID NO:18)	501	Chr11:1243966-1244466
MUC5B-Prim-5 Seq Pr.	TTGGCTAAGGTGGGAGACCT (SEQ ID NO:19)			Chr11:1241791-1241810

Table 4. SNPs identified in resequencing of the *MUC5B* promoter

Hg19 Position	SNP Name	Base Change	SEQ ID NO:	Flanking Sequence
240338	rs2672792	T/C	20	GTCACCTGCCCAGGTCCCCGAGGCC[T/C]GGAACACCTTCCTGCTGGGCCCACC
240485	rs72636989	G/A	21	CCACCCCAGGAGTTGGGGGGCCCCCGT[G/A]CCAGGGAGCAGGAGGCTGCCGAGG
240925	Muc5B-Prm1	C/T	22	GTGGCCCTGATCACTGGTGCCTGGA[C/T]GGCCTCTGAAGGGGTCTGTGGGGTC
241005	rs2672794	C/T	23	AACCCCCCTCGGGTTCTGTGTGGTC[C/T]AGGCCGCCCCTTTGTCTCCACTGCC
241221	rs35705950	G/T	24	TTTCTTCCTTTATCTTCTGTTTTCAGC[G/T]CCTTCAACTGTGAAGAAGTGA
241361	MUC5B-Prm2	A/G	25	TGCCCCGGACCCAGCCCAGTTCCCA[A/G]TGGGCCCTCTGCCCGGGGAGGTGC
241762	MUC5B-Prm3	C/T	26	GGTGGGCATCGGCTTGTGAGCTGGAGCCG[C/T]GGGCAGGGAGGGGGGATGTCACGAG
241821	rs11042491	G/A	27	GGCTAAGGTGGGAGACCTGGGCGGGTGC[G/A]TCGGGGGGACGTCTGCAGCAGAGGC
241848	rs2735726	T/C	28	TGCGTCGGGGGGACGTCTGCAGCAGAGGCC[T/C]GGGCAGCAGGCACACCCCTCCTGCCAG
241993	MUC5B-Prm4	G/A	29	GGGGCCTGGGTGCAGTCACAGCCAC[G/A]AGCCCAGGGGTGGGGACTCTGGCC
242092	MUC5B-Prm5	C/T	30	CCCCTCCCACCGTGCCGTGCTGCAG[C/T]GGGTCTACCGGCCTGGATGTGAAA
242101	MUC5B-Prm6	C/T	31	CCGTGCCGTGCTGCAGCGGGTCTAC[C/T]GGCCTGGATGTGAAAGAGAGCTTG
242227	rs11042646	C/T	32	AGTCCCGGAAGTGAGCGGGGAGCTA[C/T]GCTGAGATCTGGGAGACCCCCTGC
242244	rs55974837	C/T	33	GGGAGCTACGCTGAGATCTGGGAGA[C/T]CCCCTGCCCCCACCCAGGTACAGG
242250	rs35619543	G/T	34	TACGCTGAGATCTGGGAGACCCCCT[G/T]CCCCCACCCAGGTACAGGGCCAGG
242299	rs12804004	T/G	35	GCAGAAGCCCGAGGTGTGCCCTGAG[T/G]TAAAGAAACCGTCACAAAGAACAA
242472	rs56031419	G/A	36	TGTCTCCGCCCTCCATCTCCAGAAC[G/A]TTCTCACATTCCCAAGCTGAAACC
242508	rs868902	C/A	37	CCCAAGCTGAAACCCTGTCCCCATG[C/A]AACACCAGCTCACCATCCCCTCTGCC
242567	MUC5B-Prm7	C/T	38	GGCGCCCACCGTCCACACTCCGTCT[C/T]TGCGGGTTTCATGACTCCAGGGGCAG
242599	MUC5B-Prm8	G/A	39	TTTCATGACTCCAGGGGCAGCACAC[G/A]AGTGGCCCCTCCTGCCTTTGTCCTC
242607	MUC5B-Prm9	C/T	40	CTCCAGGGGCAGCACACGAGTGGCC[C/T]CTCCTGCCTTTGTCCTCTGTGTCCA
242690	rs868903	C/T	41	CCCCCATGGAGCAGCCTGGGCCAGCC[C/T]CTCCTTTTCACGGCTGAACCGTAT
242910	MUC5B-Prm10	G/A	42	ACCCCCACCAGCAGGGCACAGGGCTCC[G/A]GGTCCCCACGTCTCTGCCAACACTT
242977	MUC5B-Prm11	G/A	43	CTTGATCCCCGCCATCCTATTGAGC[G/A]TGAGACAGGTCAGAAGCTTTGAAG
243218	MUC5B-Prm12	G/A	44	GTCTGCGCCACGGAGCATTCAGGAC[G/A]CTGGTGACCAGGGAGCCAGGAGGT
243378	rs885455	A/G	45	CGTCAAGGAGGTTTACCACATAGCCCCC[A/G]GGAAGCCCACCCGACACCAGCCGGA
243391	rs885454	G/A	46	TTTACCACATAGCCCCCRGGAAGCCCACCC[G/A]ACACCAGCCGGAGGTGCTAGGCTTC
243409	MUC5B-Prm13	T/C	47	CCCACCCGACACCAGCCGGAGGTGC[T/C]AGGCTTCTGCGGCTCCCACCTGGG
243911	MUC5B-Prm14	G/A	48	GGACCCATGGTCAGTGGCTGGGGGT[G/A]CTGCCCAGAGGCTGGGATTCCCTTC
244060	rs7115457	G/A	49	GCCATCTAGGACGGGTGCCAGGTGG[G/A]GTAGGCCCTTCTCTCCCTTCCGATT
244080	rs7118568	C/G	50	GGTGGGGTAGGCCCTTCTCTCCCTT[C/G]CGATTCTCAGAAGCTGCTGGGGGTG
244197	rs56235854	G/A	51	AGCCCCTCCCCGAGAGCAAACACAC[G/A]TGGCTGGAGCGGGGAAGAGCATGGTGC
244219	rs2735738	T/C	52	CACGTGGCTGGAGCGGGGAAGAGCA[T/C]GGTGCCCTGCGTGGCCTGGCCTGGC
244438	MUC5B-Prm15	C/T	53	GCCGCAGGCAGGTAAGAGCCCCCCA[C/T]TCCGCCCCCTCTCGATGCTGTCTT

[0162] Next, the relationship between the rs35705950 SNP and the 18 other SNPs significantly associated with IIP were analyzed. Testing pairwise LD between these SNPs by the r2 statistic, 10 of the 18 SNPs were found to exhibit low level LD (r2=0.15-0.27) with rs35705950 among IPF cases, suggesting the significance of these SNPs is due to LD with rs35705950 (Fig. 3). Using genotypic logistic regression models to adjust for rs35705950 effects, we observed that the coefficients and corresponding P values were substantially reduced for all 18 SNPs which were previously associated with FIP and/or IPF (Table 5). After controlling for rs35705950, only one SNP retained nominal significance for IPF (rs41480348, P=0.04). It was demonstrated that the significance of the rs35705950 SNP was largely unaffected by adjustment for any of the 18 SNPs tested (P value for all SNP models was less than 1.7x10-9 for FIP and 1.1x10-24 for IPF; Table 5). These results demonstrate a strong independent effect of the rs3570590 SNP on both FIP and IPF.

Table 5. Genotypic logistic regression models for the 19 significant SNPs in the screen of lung-expressed gel-forming mucins alone, and after adjusting for rs35705950, in patients with IPF or FIP.

Model #	SNP	IPF Single SNP Model		IPF rs35705950		FIP Single SNP Model		FIP rs35705950	
		Odds Ratio (95% C.I)	P Value	Odds Ratio (95% C.I)	P Value	Odds Ratio (95% C.I)	P Value	Odds Ratio (95% C.I)	P Value
1	rs10902081	0.7 (0.5-0.8)	4.3×10^{-4}	0.9 (0.7-1.2)	0.429	0.6 (0.4-0.9)	0.011	0.8 (0.5-1.2)	0.25
	rs35705950	x	x	8.3 (5.7-11.9)	1.5×10^{-29}	x	x	5.9 (3.5-10.1)	6.6×10^{-11}
2	rs7127117	1.6 (1.3-2.0)	6.9×10^{-5}	1.1 (0.8-1.4)	0.509	1.0 (0.7-1.5)	0.826	0.7 (0.4-1.1)	0.094
	rs35705950	x	x	7.9 (5.4-11.6)	1.3×10^{-25}	x	x	6.3 (3.5-11.4)	7.3×10^{-10}
3	rs41453346	2.8 (1.6-5.2)	0.001	1.1 (0.6-2.2)	0.72	1.9 (0.8-4.3)	0.124	1.2 (0.5-3.0)	0.653
	rs35705950	x	x	8.1 (5.6-11.8)	2.7×10^{-28}	x	x	6.1 (3.6-10.3)	1.2×10^{-11}
4	rs41480348	0.5 (0.4-0.8)	0.001	0.6 (0.4-1.0)	0.04	0.7 (0.4-1.2)	0.188	0.9 (0.5-1.7)	0.75
	rs35705950	x	x	7.9 (5.5-11.3)	2.1×10^{-29}	x	x	6.1 (3.6-10.2)	1.0×10^{-11}
5	rs7934606	1.7 (1.4-2.2)	3.8×10^{-6}	1.0 (0.7-1.3)	0.876	1.4 (1.0-2.0)	0.055	0.9 (0.6-1.3)	0.473
	rs35705950	x	x	8.7 (5.8-12.9)	1.4×10^{-26}	x	x	6.7 (3.8-11.9)	7.5×10^{-11}
6	rs10902089	1.5 (1.2-1.9)	2.9×10^{-4}	0.9 (0.7-1.2)	0.69	1.5 (1.0-2.1)	0.031	1.0 (0.7-1.6)	0.813
	rs35705950	x	x	8.3 (5.6-12.2)	1.3×10^{-26}	x	x	6.1 (3.6-10.5)	6.2×10^{-11}
7	rs9667239	1.9 (1.4-2.7)	5.6×10^{-5}	0.8 (0.5-1.2)	0.3	2.2 (1.4-3.6)	0.001	1.1 (0.6-2.0)	0.668
	rs35705950	x	x	8.9 (6.0-13.3)	8.2×10^{-27}	x	x	5.8 (3.3-10.2)	6.0×10^{-10}
8	rs55846509	3.6 (1.7-7.3)	0.001	1.0 (0.5-2.3)	0.96	1.7 (0.6-5.1)	0.32	0.8 (0.3-2.5)	0.706
	rs35705950	x	x	8.3 (5.7-12.1)	2.7×10^{-28}	x	x	6.4 (3.8-10.7)	4.8×10^{-12}
9	rs28403537	4.6 (2.8-7.6)	3.2×10^{-9}	1.5 (0.8-2.6)	0.2	2.7 (1.3-5.3)	0.006	0.8 (0.3-1.8)	0.53
	rs35705950	x	x	7.6 (5.2-11.2)	1.1×10^{-24}	x	x	6.7 (3.8-11.8)	4.7×10^{-11}
10	MUC5AC-025447	1.6 (1.2-2.2)	0.003	1.1 (0.8-1.6)	0.49	1.6 (1.0-2.5)	0.053	1.4 (0.8-2.4)	0.19
	RS35705950	x	x	7.7 (5.3-11.2)	3.1×10^{-27}	x	x	6.0 (3.5-10.3)	4.7×10^{-11}
11	rs35288961	2.0 (1.5-2.6)	3.7×10^{-6}	1.1 (0.8-1.5)	0.58	2.2 (1.4-3.5)	3.2×10^{-4}	1.3 (0.7-2.1)	0.384
	rs35705950	x	x	7.9 (5.4-11.5)	6.6×10^{-27}	x	x	5.7 (3.3-10.0)	1.3×10^{-9}
12	rs35671223	1.5 (1.2-1.9)	0.001	0.9 (0.7-1.2)	0.46	1.4 (1.0-2.0)	0.05	0.9 (0.6-1.4)	0.61
	rs35705950	x	x	8.5 (5.8-12.4)	1.1×10^{-28}	x	x	6.3 (3.6-10.9)	5.4×10^{-11}
13	rs28654232	0.6 (0.5-0.8)	1.1×10^{-4}	0.9 (0.7-1.1)	0.29	0.6 (0.4-0.9)	0.009	0.7 (0.5-1.1)	0.167
	rs35705950	x	x	8.0 (5.5-11.5)	5.7×10^{-29}	x	x	5.7 (3.4-9.6)	5.8×10^{-11}
14	rs34595903	0.5 (0.4-0.7)	2.4×10^{-6}	0.8 (0.6-1.1)	0.116	0.5 (0.3-0.7)	0.001	0.6 (0.4-1.0)	0.041
	rs35705950	x	x	7.4 (5.1-10.8)	7.0×10^{-26}	x	x	5.1 (3.0-8.6)	1.7×10^{-9}
15	rs2672794	0.5 (0.4-0.7)	1.9×10^{-7}	0.9 (0.7-1.2)	0.442	0.5 (0.3-0.8)	0.001	0.7 (0.4-1.1)	0.152
	rs35705950	x	x	8.0 (5.5-11.6)	2.5×10^{-27}	x	x	5.5 (3.2-9.3)	3.2×10^{-10}
16	rs35619543	2.1 (1.6-2.8)	1.5×10^{-8}	1.3 (0.9-1.7)	0.145	2.4 (1.6-3.6)	3.3 x10-5	1.3 (0.8-2.1)	0.296
	rs35705950	x	x	7.6 (5.2-11.2)	7.0×10^{-25}	x	x	6.1 (3.4-10.9)	6.8×10^{-10}
17	rs12804004	0.6 (0.5-0.8)	1.2×10^{-4}	0.8 (0.6-1.0)	0.07	0.6 (0.4-0.9)	0.019	0.7 (0.5-1.1)	0.159
	rs35705950	x	x	7.9 (5.5-11.3)	6.4×10^{-29}	x	x	5.9 (3.5-10.0)	3.6×10^{-11}
18	rs868903	1.6 (1.3-2.1)	2.8×10^{-5}	1.0 (0.8-1.4)	0.753	1.8 (1.3-2.6)	0.001	1.4 (0.9-2.0)	0.145
	rs35705950	x	x	7.8 (5.3-11.5)	8.6×10^{-26}	x	x	5.6 (3.2-9.6)	4.4×10^{-10}

Example 2: Single nucleotide polymorphism rs35705950 results in increased expression of *MUC5B* gene

[0163] The wildtype G allele of the rs35705950 SNP is conserved across primate species. The SNP is directly 5' to a highly conserved region across vertebrate species, and is in the middle of sequence predicted to be involved in *MUC5B* gene regulation. A bioinformatic analysis of the effect of the rs35705950 SNP predicts a disruption of an E2F binding site and creation of at least two new binding sites (*e.g.* HOX9 and PAX-2).

[0164] Based on these analyses, the effect of rs35705950 was examined on *MUC5B* gene expression. In lung tissue from 33 subjects with IPF and 47 unaffected subjects, quantitative RT-PCR revealed that *MUC5B* gene expression was upregulated 14.1-fold among IPF subjects compared to unaffected subjects (P=0.0001, Fig. 4A). A 37.4-fold increase in *MUC5B* expression was observed among unaffected subjects carrying at least one copy of the variant allele compared to homozygous wildtype subjects (P=0.0003, Fig. 4B). In contrast, no significant difference in *MUC5B* gene expression was observed among the IPF subjects with at least one variant allele of rs35705950 (Fig. 4C). Smoking, a potential confounder of *MUC5B* expression, appeared to have little effect on the association between the rs35705950 variant allele and *MUC5B* expression among either unaffected or IPF affected subjects (Figs. 4B and 4C).

[0165] MUC5B immunohistochemical staining in lung tissue showed cytoplasmic staining in secretory columnar cells of the bronchi and larger proximal bronchioles (>200 μm) in IPF cases and controls (Fig. 5A). In subjects with IPF, regions of dense accumulation of MUC5B were observed in areas of microscopic honeycombing and involved patchy staining of the metaplastic epithelia lining the honeycomb cysts (Fig. 5B), as well as the mucous plugs within the cysts (Fig. 5C). No obvious differences were observed in MUC5B staining characteristics in IPF cases with the MUC5B promoter polymorphism.

[0166] IPF subjects have significantly more MUC5B lung gene expression than controls, and MUC5B protein is expressed in pathologic lesions of IPF. The present results show that the risk of developing FIP or IPF is substantially correlated with the re35705950 promoter polymorphism, which causes increased MUC5B expression. In aggregate, the data show that MUC5B expression in the lung plays a role in the pathogenesis of pulmonary disease.

[0167] Based on the relationship between the SNP and excess production of MUC5B, too much MUC5B can impair mucosal host defense to excessive lung injury from inhaled substances, and, over time, lead to the development of IIP. In addition to the *MUC5B* promoter SNP, common exposures and basic biological processes can influence either the expression or clearance of MUC5B. For instance, MUC5B expression can be enhanced in the lung by cigarette smoke, acrolein, oxidative stress, IL-6, IL-8, IL-13, IL-17, 17β-estradiol, extracellular nucleotides, or epigenetic changes that alter DNA methylation or chromatin structure. Moreover, clearance of lung mucus is dependent on effective ciliary motion, adequate hydration of the periciliary liquid layer, and an intact cough. Regardless of the cause, the present results indicate that excess MUC5B can compromise mucosal host defense, reducing lung clearance of inhaled particles, dissolved chemicals, and microorganisms. Given the importance of environmental exposures, such as asbestos, silica, and other pollutants in the development of other forms of interstitial lung disease, it is logical to speculate that common inhaled particles, such as those associated with cigarette smoke or air pollution, might lead to or contribute to exaggerated interstitial injury in individuals who have defects in mucosal host defense.

[0168] In addition, excess MUC5B in the respiratory bronchioles can interfere with alveolar repair. Alveolar injury can lead to collapse of bronchoalveolar units and this focal lung injury is repaired through re-epithelialization of the alveolus by type II alveolar epithelial cells. Thus, MUC5B can impede alveolar repair by either interfering with the interaction between the type II alveolar epithelial cells and the underlying matrix, or by interfering with the surface tension properties of surfactant. Either failure to re-epithelialize the basal lamina of the alveolus or suboptimal surfactant biology could enhance ongoing collapse and fibrosis of adjacent bronchoalveolar units, and eventually result in IIP.

[0169] Lesions of IPF are spatially heterogeneous, suggesting that IPF is multifocal, originating in individual bronchoalveolar units. Since SNP rs35705950 occurs in the promoter region of *MUC5B* and is predicted to disrupt transcription factor binding sites, ectopic production of MUC5B in cells or locations that cause injury to the bronchoalveolar unit can be a causative agent. Unscreened genetic variants (especially in the inaccessible repetitive mucin regions) may be in linkage disequilibrium with the *MUC5B* promoter SNP and affect the function of other lung mucins.

[0170] The present observations provide a novel clinical approach to pulmonary disorders such as IIP. Invoking secreted airway mucins in the pathogenesis of pulmonary fibrosis suggests that the airspace plays a role in the pathogenesis of IIP. While the SNP (rs35705950) in the *MUC5B* promoter can be used to identify individuals at risk for developing IIP, the observation that mucin biology is be important in the etiology of IIP reorients the focus of pathogenic and therapeutic studies in interstitial lung disease to lung mucins and the airspace. Moreover, the genetic causes of IIP (*e.g.*, *MUC5B*, surfactant protein C, surfactant protein A2, and the two telomerase genes TERC and TERT) provide insight into the unique clinical manifestations of this complex disease process, and consequently, lead to earlier detection, more predictable prognosis, and personalized therapeutic strategies.

Example 3: Genetic variant *MUC5B* associated with attenuated form of pulmonary disease

[0171] The data described herein demonstrate that the genetic variant *MUC5B* rs35705950 is associated with development of pulmonary disease. We next examined whether rs35705950 genetic variant is associated with disease severity and prognosis. We found that homozygous wildtype subjects (GG), *i.e.*, those having normal *MUC5B* gene sequence, displayed a steeper decline in forced vital capacity (FVC) over time as compared to subjects with at least one T allele (P=0.0006). Essentially, while FVC declines for both groups, the decline is more gradual in those carrying the G→ T polymorphism. For GG subjects, the FVC absolute value fell from about 3.4 liters to about 3.1 liters over years 1-3 of the study. For GT and TT subjects, the FVC absolute value still fell, but started at over 3.5 liters and fell to about 3.4 liters over years 1-3 of the study.

[0172] Additionally, we observed an association between death with subjects having at least one T allele having a lower mortality (OR(95%CI) = .37(.20-.67); p=0.001) after adjusting for gender, history of smoking and DLCO (diffusion lung capacity for CO2). We also observed an association with time to death and the T allele; Hazard ratio (95% CI) = .50(.30-.83) p=.0069 after adjustment for gender, history of smoking and DLCO. These results suggest that in addition to being a strong risk factor for pulmonary disease development, the rs35705950 SNP can indicate a less severe prognosis for the pulmonary disease.

WHAT IS CLAIMED IS:

1. A method of determining whether a subject has or is at risk of developing a pulmonary disease, said method comprising detecting whether a genome of a subject comprises a genetic variant *MUC5B* gene, wherein the presence of said genetic variant *MUC5B* gene indicates said subject has or is at risk of developing a pulmonary disease.

2. The method of claim 1, wherein the pulmonary disease is an interstitial lung disease.

3. The method of claim 2, wherein the interstitial lung disease is a fibrotic interstitial lung disease.

4. The method of claim 3, wherein said fibrotic interstitial lung disease is idiopathic pulmonary fibrosis or familial interstitial pneumonia.

5. The method of claim 1, wherein the presence of said genetic variant *MUC5B* gene in said subject results in overexpression of MUC5B RNA or MUC5B protein.

6. The method of claim 1, further comprising administering an pulmonary disease treatment.

7. The method of claim 1, wherein said subject is homozygous for said genetic variant *MUC5B* gene.

8. The method of claim 1, wherein said subject is heterozygous for said genetic variant *MUC5B* gene.

9. The method of claim 1, wherein said genetic variant *MUC5B* gene is a genetic variant regulatory region *MUC5B* gene.

10. The method of claim 1, wherein said genetic variant *MUC5B* gene is a genetic variant promoter *MUC5B* gene.

11. The method of any one of claim 1 - 10, wherein said genetic variant *MUC5B* gene has a single nucleotide polymorphism.

12. The method of claim 11, wherein the single nucleotide polymorphism is rs2672792, rs72636989, *MUC5B*-Prm1, rs2672794, rs35705950, *MUC5B*-Prm2, rs11042491, rs2735726, rs868902, *MUC5B*-Prm3, *MUC5B*-Prm4, *MUC5B*-Prm5, rs868903, *MUC5B*-Prm6, rs885455, rs885454, *MUC5B*-Prm7, rs7115457, rs7118568 rs56235854 or rs2735738.

13. The method of claim 11, wherein said genetic variant *MUC5B* gene is rs35705950.

14. The method of claim 11, wherein said detecting comprises determining the sequence of at least 8 contiguous nucleotides of SEQ ID NO:24, wherein said at least 8 contiguous nucleotides include the nucleotide at position 28 of SEQ ID NO:24.

15. The method of claim 11, wherein said detecting comprises determining the identity of the nucleotide at position 28 of SEQ ID NO:24.

16. The method of claim 11, wherein said genetic variant *MUC5B* gene comprises a first single nucleotide polymorphism and a second single nucleotide polymorphism.

17. The method of claim 16, where said first single nucleotide polymorphism is present within a first DNA strand and said second single nucleotide polymorphism is present within a second DNA strand.

18. The method of claim 16, where said first and second single nucleotide polymorphisms are present within a first DNA strand.

19. The method of any one of claims 1-18, wherein said method further comprises detecting whether said subject expresses an elevated MUC5B RNA level or an elevated MUC5B protein level relative to a standard control, wherein the presence of said elevated MUC5B RNA level or said elevated MUC5B protein level indicates said subject has or is at risk of developing a pulmonary disease.

20. The method claim 19, wherein said elevated MUC5B protein level or said elevated MUC5B RNA level is detected from a pulmonary tissue or bodily fluid of said subject.

21. The method claim 20, wherein said bodily fluid is whole blood, plasma, serum, urine, sputum, saliva, bronchoalveolar lavage sample, or exhaled breath condensate.

22. The method of claim 21, wherein the method further comprises isolating exosomes or subcellular vesicles from said bodily fluid.

23. The method of claim 21, wherein said bodily fluid is plasma, whole blood or serum.

24. A method of determining whether a subject has or is at risk of developing a pulmonary disease, said method comprising detecting whether a subject expresses an elevated MUC5B RNA level or an elevated MUC5B protein level relative to a standard control, wherein the presence of said elevated MUC5B RNA level or said elevated MUC5B protein level indicates said subject has or is at risk of developing a pulmonary disease.

25. The method of claim 24, wherein said pulmonary disease is an interstitial lung disease.

26. The method of claim 25, wherein said interstitial lung disease is a fibrotic interstitial lung disease.

27. The method of claim 26, wherein said fibrotic interstitial lung disease is idiopathic pulmonary fibrosis or familial interstitial pneumonia.

28. The method of claim 24, further comprising administering a treatment for pulmonary disease or interstitial lung disease.

29. The method of claim 28, further comprising determining a second MUC5B RNA level or second MUC5B protein level after administering said treatment and comparing said second level to the level observed before administering said treatment.

30. The method of claim 24, wherein said elevated MUC5B protein level or said elevated MUC5B RNA level is detected in a pulmonary tissue or bodily fluid of said subject.

31. The method claim 30, wherein said bodily fluid is whole blood, plasma, serum, urine, sputum, saliva, a bronchoalveolar lavage sample, or exhaled breath condensate.

32. The method of claim 31, wherein the method further comprises isolating exosomes or subcellular vesicles from said bodily fluid.

33. The method of claim 31, wherein said bodily fluid is plasma, whole blood or serum.

34. A method of prognosing a pulmonary disease in a patient, said method comprising detecting whether a genome of a subject comprises a genetic variant *MUC5B* gene, wherein the presence of said genetic variant *MUC5B* gene indicates an attenuated form of said pulmonary disease in said patient relative to the absence of said genetic variant *MUC5B* gene.

35. The method of claim 34, wherein said pulmonary disease is a fibrotic interstitial lung disease.

36. The method of claim 35, wherein said fibrotic interstitial lung disease is idiopathic pulmonary fibrosis or familial interstitial pneumonia.

37. The method of claim 34, wherein the presence of said genetic variant *MUC5B* gene in said subject results in overexpression of MUC5B RNA or MUC5B protein.

38. The method of claim 34, further comprising administering a pulmonary disease treatment or an interstitial lung disease treatment.

39. The method of claim 34, wherein said subject is homozygous for said genetic variant *MUC5B* gene.

40. The method of claim 34, wherein said subject is heterozygous for said genetic variant *MUC5B* gene.

41. The method of claim 34, wherein said subject is negative for said genetic variant *MUC5B* gene.

42. The method of claim 34, wherein said genetic variant *MUC5B* gene is a genetic variant regulatory region *MUC5B* gene.

43. The method of claim 34, wherein said genetic variant *MUC5B* gene is a genetic variant promoter *MUC5B* gene.

44. The method of one of claims 34 - 43, wherein said genetic variant *MUC5B* gene has a single nucleotide polymorphism.

45. The method of claim 44, wherein the single nucleotide polymorphism is rs2672792, rs72636989, *MUC5B*-Prm1, rs2672794, rs35705950, *MUC5B*-Prm2, rs11042491, rs2735726, rs868902, *MUC5B*-Prm3, *MUC5B*-Prm4, *MUC5B*-Prm5, rs868903, *MUC5B*-Prm6, rs885455, rs885454, *MUC5B*-Prm7, rs7115457, rs7118568 rs56235854 or rs2735738.

46. The method of claim 44, wherein said genetic variant *MUC5B* gene is rs35705950.

47. The method of claim 44, wherein said detecting comprises determining the sequence of at least 8 contiguous nucleotides of SEQ ID NO:24, wherein said at least 8 contiguous nucleotides include the nucleotide at position 28 of SEQ ID NO:24.

48. The method of claim 44, wherein said detecting comprises determining the identity of the nucleotide at position 28 of SEQ ID NO:24.

49. The method of claim 44, wherein said genetic variant *MUC5B* gene comprises a first single nucleotide polymorphism and a second single nucleotide polymorphism.

50. The method of claim 49, where said first single nucleotide polymorphism is present within a first DNA strand and said second single nucleotide polymorphism is present within a second DNA strand.

51. The method of claim 49, where said first and second single nucleotide polymorphisms are present within a first DNA strand.

52. The method of any one of claims 34-51, wherein said method further comprises detecting whether said subject expresses an elevated MUC5B RNA level or an elevated MUC5B protein level relative to a standard control, wherein the presence of said elevated MUC5B RNA level or said elevated MUC5B protein level indicates an attenuated course of said interstitial lung disease in said patient.

53. The method claim 52, wherein said elevated MUC5B protein level or said elevated MUC5B RNA level is detected from a pulmonary tissue or bodily fluid of said subject.

54. The method claim 53, wherein said bodily fluid is whole blood, plasma, serum, urine, sputum, saliva, bronchoalveolar lavage sample, or exhaled breath condensate.

55. The method of claim 54, wherein method further comprises isolating exosomes or subcellular vesicles from said bodily fluid.

56. The method of claim 54, wherein said bodily fluid is plasma, whole blood or serum.

57. A method of treating a pulmonary disease in a subject in need thereof, said method comprising

administering to said patient an effective amount of a nucleic acid or an antibody,

wherein said nucleic acid is capable of hybridizing to a *MUC5B* gene or a MUC5B RNA thereby decreasing levels of MUC5B protein in said subject, and

wherein said antibody is capable of binding to a MUC5B protein thereby decreasing activity of said MUC5B protein, thereby treating the pulmonary disease in the subject.

58. The method of claim 57, wherein said pulmonary disease is interstitial lung disease.

59. The method of claim 58, wherein said interstitial lung disease is a fibrotic interstitial lung disease.

60. The method of claim 59, wherein said fibrotic interstitial lung disease is idiopathic pulmonary fibrosis or familial interstitial pneumonia.

61. The method of claim 57, further comprising administering to said patient an effective amount of a mucolytic drug.

62. A method of treating a pulmonary disease in a subject comprising

detecting whether the genome of the subject comprises a genetic variant *MUC5B* gene; and

administering to the subject a pulmonary disease treatment or a MUC5B antagonist, thereby treating the pulmonary disease in the subject.

63. The method of claim 62, further comprising determining whether said subject expresses an elevated MUC5B RNA level or an elevated MUC5B protein level relative to a standard control before administering the pulmonary disease treatment or MUC5B antagonist.

64. The method of claim 62, wherein said genetic variant MUC5B is rs35705950.

65. A method of treating a pulmonary disease in a subject, comprising

determining whether a subject expresses an elevated MUC5B RNA level or an elevated MUC5B protein level relative to a standard control; and

when an elevated level of MUC5B RNA or MUC5B protein is found relative to the standard control, administering to the subject a pulmonary disease treatment or a MUC5B antagonist, thereby treating the pulmonary disease in the subject.

66. The method of any one of claim 62-65, wherein said pulmonary disease treatment is a mucolytic drug.

67. A method of identifying a candidate pulmonary disease treatment compound, said method comprising

(i) contacting a test compound with a MUC5B protein;

(ii) allowing said test compound to inhibit the activity of said MUC5B protein; and

(iii) selecting the test compound that inhibits the activity of said MUC5B protein, thereby identifying a candidate pulmonary disease treatment compound.

68. The method of claim 67, wherein said pulmonary disease is an interstitial lung disease.

69. The method of claim 68, wherein said interstitial lung disease is a fibrotic interstitial lung disease.

70. The method of claim 67, wherein said method is *in vitro*.

71. The method of claim 67, wherein said method is *in vivo*.

72. A method of identifying a candidate pulmonary disease treatment compound, said method comprising:

(i) contacting a test compound with a MUC5B secreting cell;

(ii) allowing said test compound to inhibit secretion of MUC5B protein from said MUC5B secreting cell; and

(iii) selecting the test compound that inhibits secretion of MUC5B protein from said MUC5B secreting cell,

thereby identifying a candidate pulmonary disease treatment compound.

73. The method of claim 72, wherein said pulmonary disease is an interstitial lung disease.

74. The method of claim 73, wherein said interstitial lung disease is a fibrotic interstitial lung disease.

75. The method of claim 72, wherein said MUC5B secreting cell forms part of a pulmonary tissue.

76. The method of claim 72, wherein said MUC5B secreting cell is in vitro.

77. The method of claim 75, wherein said pulmonary tissue forms part of an organism.

78. The method of claim 77, wherein said organism is a mammal.

79. The method of claim 78, wherein said organism is a human.

80. A method of determining whether a fibrotic interstitial lung disease is progressing in an interstitial lung disease patient, said method comprising:

(i) determining a first level of MUC5B RNA or first level of MUC5B protein in said patient at a first time point;

(ii) determining a second level of MUC5B RNA or second level of MUC5B protein in said patient at a second time point;

(iii) comparing the second level of MUC5B RNA to the first level of MUC5B RNA or comparing the second level of MUC5B protein to the first level of MUC5B protein, wherein if the second level of MUC5B RNA is greater than the first level of MUC5B RNA or if the second level of MUC5B protein is greater than the first level of MUC5B protein, the interstitial lung disease is progressing in the patient.

81. The method of claim 80, further comprising determining a rate of progression of said fibrotic interstitial lung disease in said patient based on said comparing.

82. The method of claim 80, wherein said determining said first level of MUC5B RNA or first level of MUC5B protein and said second level of MUC5B RNA or second level of MUC5B protein comprises normalizing said first level of MUC5B RNA or first level of MUC5B protein and said second level of MUC5B RNA or second level of MUC5B protein to a level of RNA or protein expressed from a standard gene in said interstitial lung disease patient.

83. The method of claim 82, wherein said standard gene is a GAPDH gene or a beta-actin gene.

84. A kit comprising:

(a) a MUC5B binding agent capable of binding to a substance selected from the group consisting of:

(i) a genetic variant *MUC5B* gene sequence;

(ii) a MUC5B RNA or fragment thereof; and

(iii) a MUC5B protein or fragment thereof, and

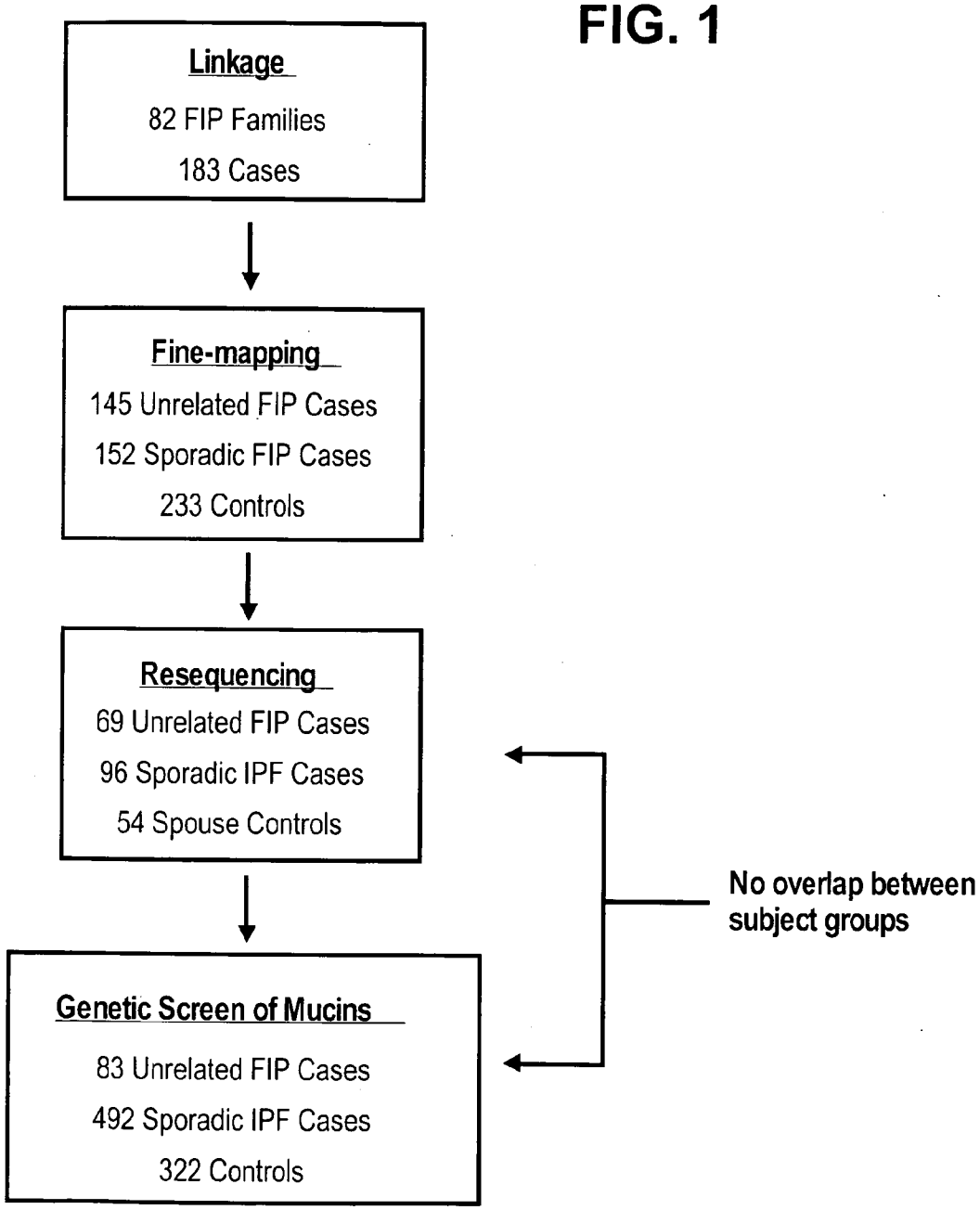
(b) a detecting reagent or a detecting apparatus capable of indicating binding of said MUC5B binding agent to said substance.

85. The kit of claim 84, further comprising

c) a sample collection device for collecting a sample from a subject.

86. The kit of claim 84, wherein said genetic variant *MUC5B* gene sequence is in the promoter of said *MUC5B* gene sequence.

FIG. 1


Linkage
82 FIP Families
183 Cases
Fine-mapping
145 Unrelated FIP Cases
152 Sporadic FIP Cases
233 Controls
Resequencing
69 Unrelated FIP Cases
96 Sporadic IPF Cases
54 Spouse Controls
Genetic Screen of Mucins
83 Unrelated FIP Cases
492 Sporadic IPF Cases
322 Controls
No overlap between subject groups

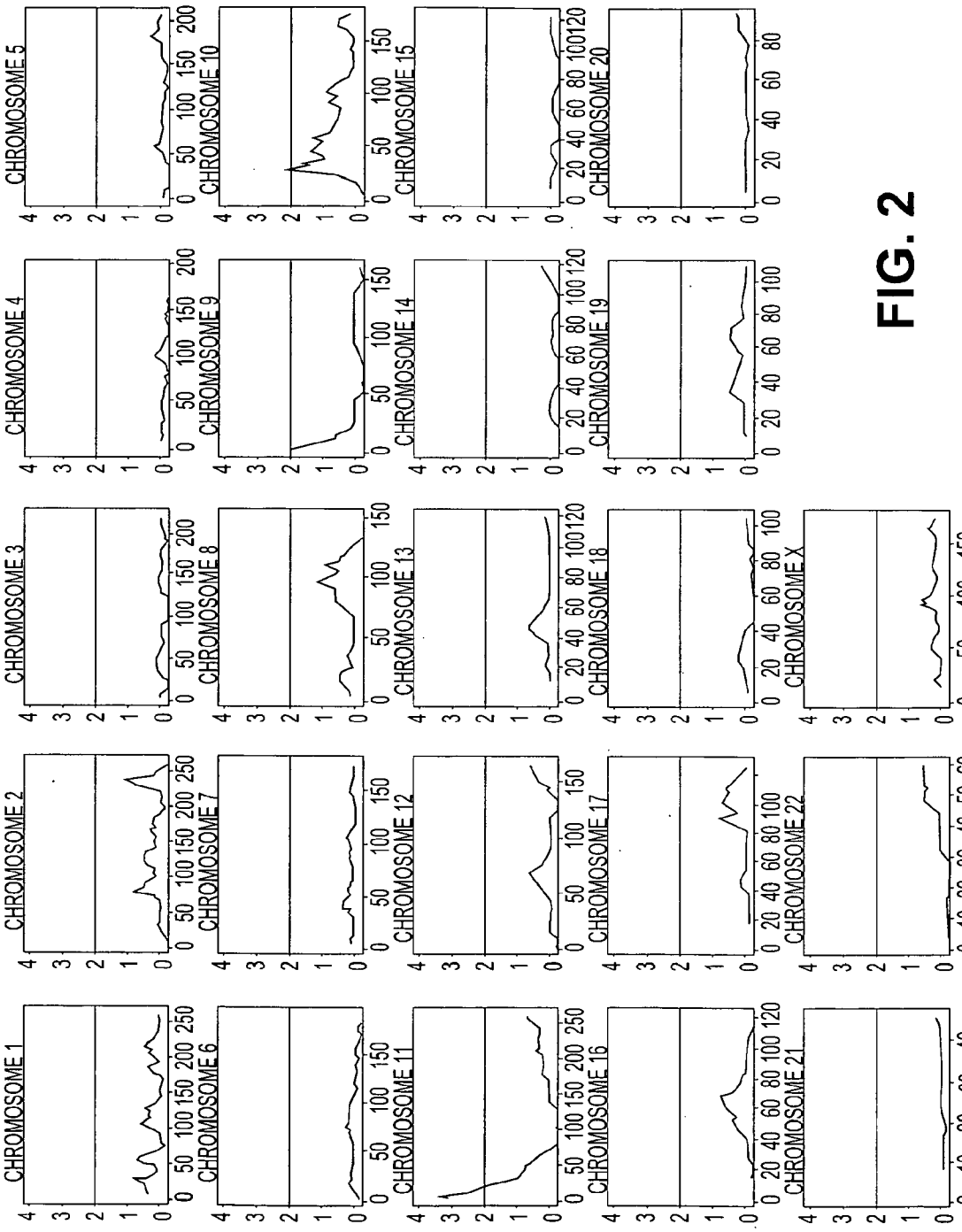

CHROMOSOME 1
CHROMOSOME 2
CHROMOSOME 3
CHROMOSOME 4
CHROMOSOME 5
CHROMOSOME 6
CHROMOSOME 7
CHROMOSOME 8
CHROMOSOME 9
CHROMOSOME 10
CHROMOSOME 11
CHROMOSOME 12
CHROMOSOME 13
CHROMOSOME 14
CHROMOSOME 15
CHROMOSOME 16
CHROMOSOME 17
CHROMOSOME 18
CHROMOSOME 19
CHROMOSOME 20
CHROMOSOME 21
CHROMOSOME 22
CHROMOSOME X

FIG. 2

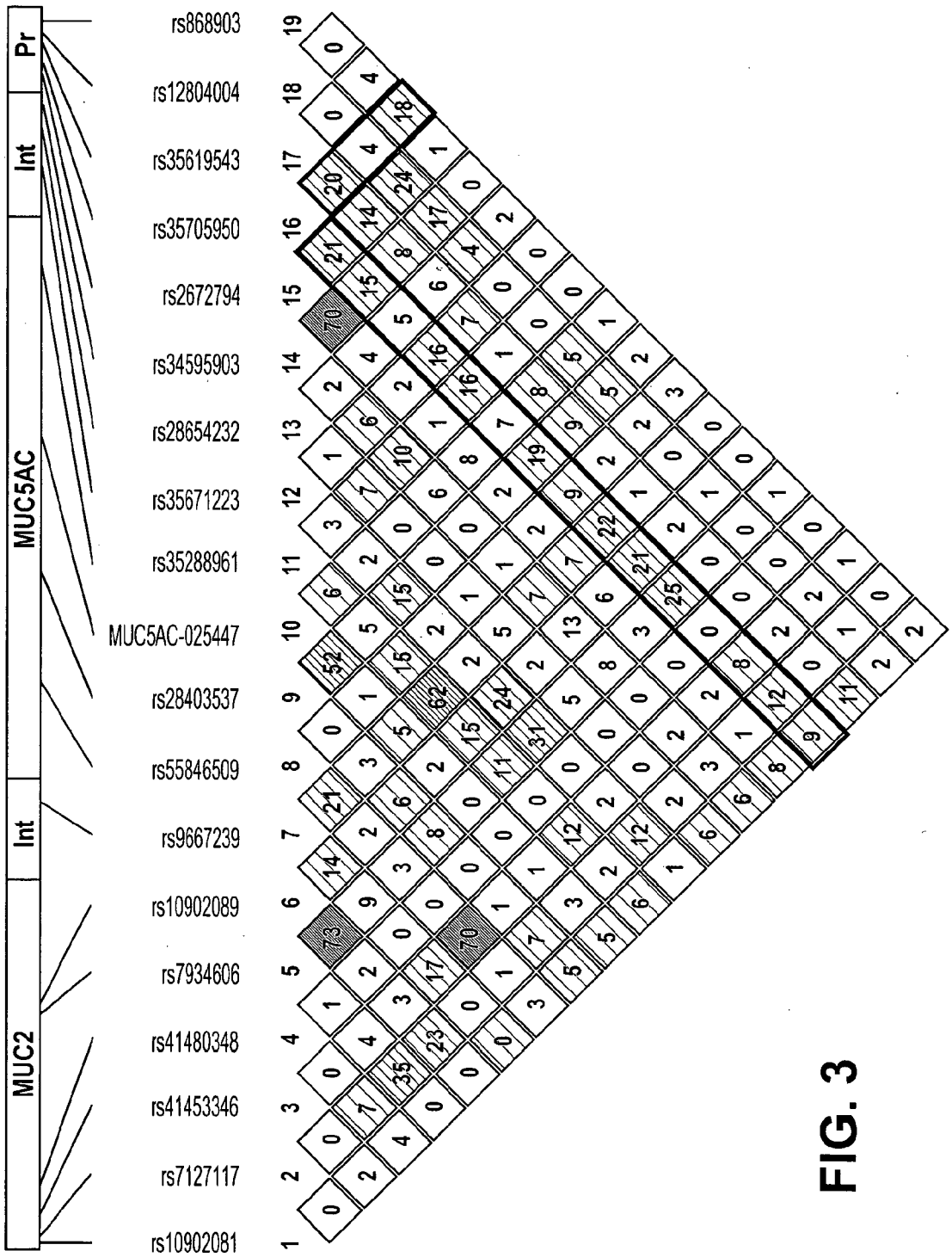

Pr
Int
MUC5AC
Int
MUC2
rs868903
rs12804004
rs35619543
rs35705950
rs2672794
rs34595903
rs28654232
rs35671223
rs35288961
MUC5AC-025447
rs28403537
rs55846509
rs9667239
rs10902089
rs7934606
rs41480348
rs41453346
rs7127117
rs10902081

FIG. 3

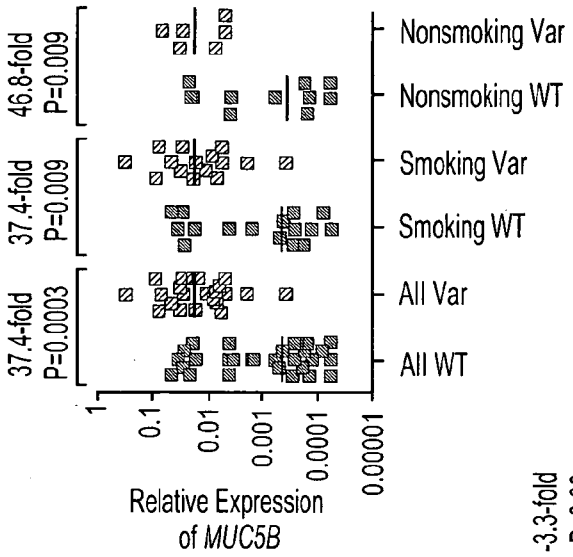


FIG. 4B

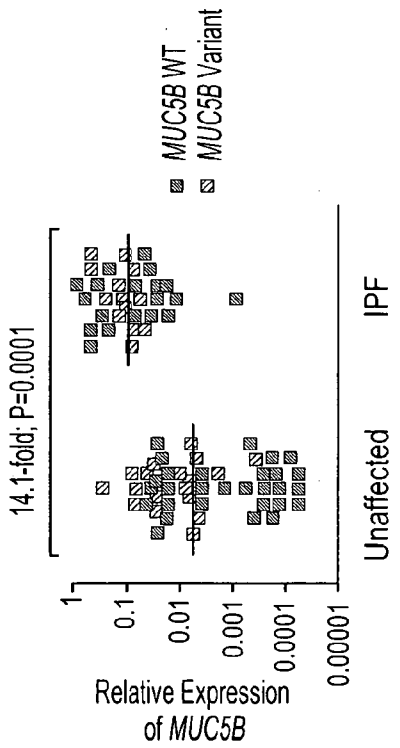


FIG. 4A

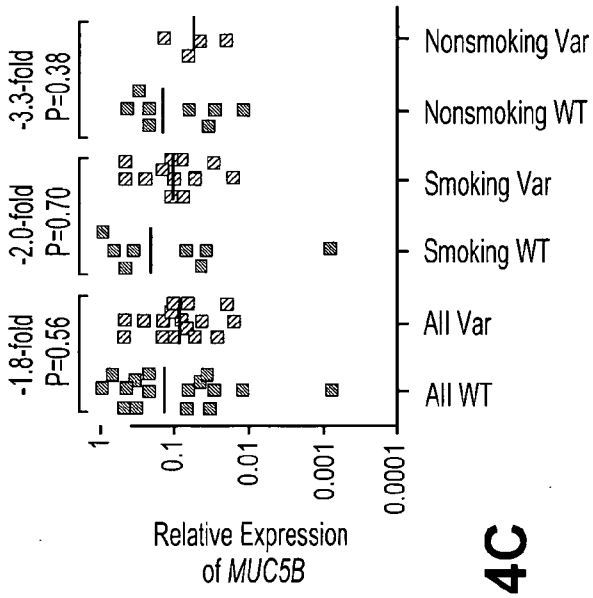


FIG. 4C

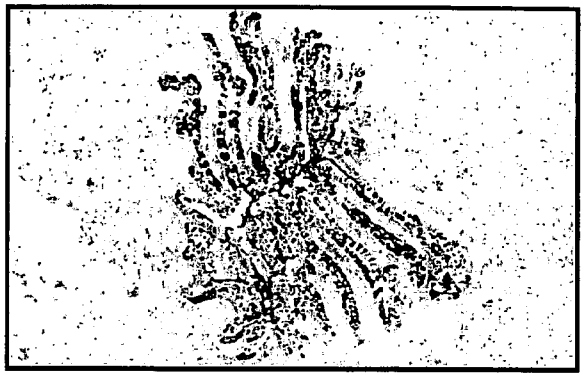


FIG. 5A

FIG. 5B

FIG. 5C

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 11/22621

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12Q 1/68; G01N 33/68 (2011.01)
USPC - 435/6; 435/7.1
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)
IPC(8): C12Q 1/68; G01N 33/68 (2011.01)
USPC: 435/6; 435/7.1

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

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
--- Please see extra sheet for continuation ----

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	KAMIO et al., Promoter Analysis and Aberrant Expression of the MUC5B Gene in Diffuse Panbronchiolitis. Amer J of Respiratory Critical Care Medicine. 1 May 2005, Vol 171, No 9, Pages 949-957, especially page 949, abstract; page 950, left col, 2nd para; page 951, Table 2; page 952, Table 3; page 952, Fig 2; page 953, Fig 4.	1, 7-11, 16, 18, 24, 30, 31 --- 2-6, 12, 13, 17, 25-29, 32-43, 44/(34-36, 38-43), 45-46, 49-51, 57-66, 84-86
Y	WO 2009/073167 A2 (SCHWARTZ et al.) 11 June 2009 (11.06.2009), pg 1, ln 5-8; pg 2, ln 10-15; pg 14, ln 3-5; pg 16, ln 30-31; pg 28, ln 5-7; pg 32, ln 27-30; pg 43, ln 5-6; pg 66, ln 14-16; pg 67, ln 7-20; pg 69, ln 16; pg 69, ln 2-6; pg 69, ln 22-29; pg 69, ln 33-34; pg 70, ln 15-20; pg 70, ln 25-30; pg 73, ln 32-34; pg 74, ln 7-16; pg 74, ln 28-31; pg 82, ln 15; pg 83, ln 22-25; pg 84, ln 33-34; pg 86, ln 3-4; pg 86, ln 4-5; pg 86, ln 5-7; pg 86, ln 15-16; pg 87, ln 3-22; pg 90, ln 23-26; pg 90, ln 28-34; pg 91, ln 5-9; pg 92, ln 28-31; pg 93, ln 8-9; pg 93, ln 8-13; pg 94, ln 14-18; pg 94, ln 1-6; pg 94, ln 22-27; pg 96, ln 15-20; Table 5.	2-4, 6, 17, 25-29, 34-43, 44/(34-36, 38-43), 45-46, 49-51, 57-86 5, 37, 44/(37), 67-83
Y --- A	US 2003/0096219 A1 (WU et al.) 22 May 2003 (22.05.2003), abstract; para [0005], [0012], [0014], [0022], [0030], [0032], [0099], [0100], [0109], [0156], [0189], [0215], [0230], SEQ ID NO: 32.	14, 15, 47, 48 12, 13, 45, 46, 64
Y	NCBI: Single Nucleotide Polymorphism. Cluster report: rs35705950; map; 2005-06-16, NCBI SNP database [online], [retrieved on 2011-03-25] Retrieved from the Internet: <URL: http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ss.cgi?subsnp_id=38666420>	

☒ Further documents are listed in the continuation of Box C. ☐

* 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	Date of mailing of the international search report
28 March 2011 (28.03.2011)	18 APR 2011
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (July 2009)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 11/22621

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.: 19-23 and 52-56
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.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 11/22621

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	KESIMER et al., Characterization of exosome-like vesicles released from human tracheobronchial ciliated epithelium: a possible role in innate defense. FASEB Journal, June 2009, Vol 23, No 6, Pages 1858-1868, especially page 1858, abstract; pg 1859, left col, 3rd para; page 1862, left col, 2nd para.	32, 33
Y	NGUYEN et al. Chronic Exposure to Beta-Blockers Attenuates Inflammation and Mucin Content in a Murine Asthma Model. Amer J of Respiratory Cell Mol Biol, 2008, Vol 38, No 3, Pages 256-262, especially page 256, abstract; page 259, right col, 1st para.	34-,43, 44/(34-36, 38-43), 45-46, 49-51
Y	US 6,339,075 B1 (KING et al.) 15 January 2002 (15.01.2002), abstract	61, 66
Y	US 2005/0125851 A1 (WHITSETT et al.) 9 June 2005 (09.06.2005), abstract; para [0009], [0018], [0213], [0225]; Fig 1.	80-83

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 11/22621

Continuation on (B): Fields Searched - Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

PubWEST (PGPB,USPT,EPAB,JPAB): mucin, MUC5B, MG1, MUC5, MUC9, MUC-5B, rs35705950, single nucleotide polymorphism, SNP, interstitial pneumonia, idiopathic, fibrosis, heterozygous, homozygous, mucolytic, rs35705950, N-acetylcysteine, N-acystelyn, erdoseine, dextran, pulmozyme

Google Scholar: N-acetylcysteine dextran, pulmozyme mucolytic MUC5B genetic variant homozygous heterozygous SNP exosomes attenuated

NCBI SNP: rs35705950

NCBI: MUC5B

esp@cenet: David Schwartz, Mucin, National Jewish Health