

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
5 October 2006 (05.10.2006)

PCT

(10) International Publication Number
WO 2006/105252 A2

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

(21) International Application Number:
PCT/US2006/011570

(22) International Filing Date: 28 March 2006 (28.03.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/666,085 28 March 2005 (28.03.2005) US

(71) Applicant (for all designated States except US): **THE REGENTS OF THE UNIVERSITY OF COLORADO** [US/US]; 201 Regent Administrative Center, 3 SYS, Boulder, CO 80309 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **GERACI, Mark, W.** [US/US]; 5093 S. Hannibal Way, Aurora, CO 80015 (US). **COLDREN, Christopher, D.** [US/US]; 2409 S. Dahlia Lane, Denver, CO 80222 (US). **GRUBER, Michael, P.** [US/US]; 9161 E. Arizona Place, Denver, CO 80247 (US). **SULLIVAN, Andrew, K.** [US/US]; 680 Fox Street, Denver, CO 80204 (US). **VOELKEL, Norbert, F.** [DE/US]; 1203 E. 3rd Avenue, Denver, CO 80218 (US).

(74) Agents: **YOERG, Darla, G.** et al.; Sheridan Ross P.C., 1560 Broadway, Suite 1200, Denver, Colorado 80202-5141 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

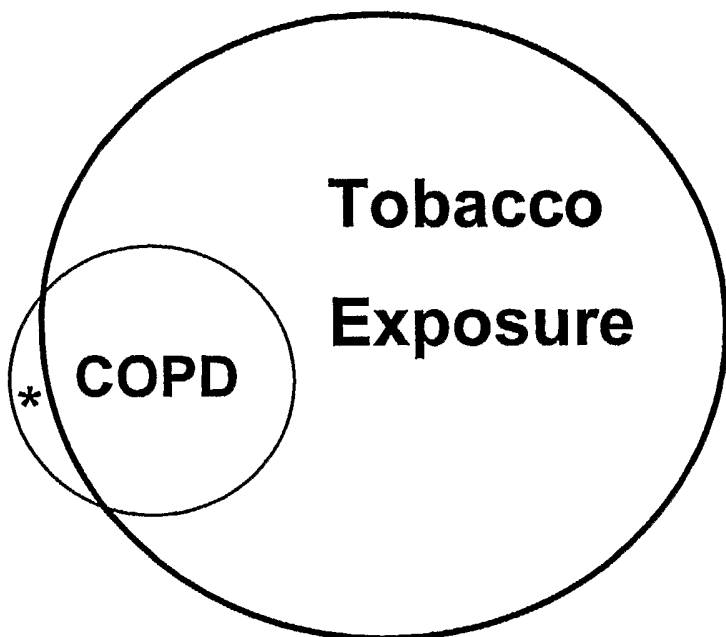
Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: DIAGNOSIS OF CHRONIC PULMONARY OBSTRUCTIVE DISEASE AND MONITORING OF THERAPY USING GENE EXPRESSION ANALYSIS OF PERIPHERAL BLOOD CELLS

(57) Abstract: Genes and proteins that are regulated differentially in individuals with chronic pulmonary obstructive disease (COPD) are provided, as well as methods for their use.



WO 2006/105252 A2

**DIAGNOSIS OF CHRONIC PULMONARY OBSTRUCTIVE DISEASE AND
MONITORING OF THERAPY USING GENE EXPRESSION ANALYSIS OF
PERIPHERAL BLOOD CELLS**

5

Background of the Invention

Chronic obstructive pulmonary disease (COPD) is a global epidemic. Worldwide, COPD is predicted to become the third most common cause of death and the fifth most common cause of disability by 2020¹. In the United States, COPD is the fourth leading cause of death and is projected to become the third leading cause of death for both males and females by the year 2020². Estimates suggest that approximately 14 million people in the United States have COPD and the prevalence continues to rise³. Additionally, COPD is an increasingly common cause of chronic morbidity, lost economic productivity, and consumption of health care resources². COPD is caused by long-term exposure to inhaled noxious gases and particles; most notably, tobacco smoke accounts for more than 90% of COPD cases in developed countries. Despite the strong causal link between tobacco smoke and COPD, only 10-20% of cigarette smokers develop clinically evident COPD (Fig. 1)^{3,4}.

Epidemiologic studies demonstrate that the progressive airflow limitation in COPD patients is due to an accelerated decline in lung function (measured as the forced expiratory volume in one second, FEV1) from the normal rate in adults of approximately 30 ml per year to 60 ml per year^{14,15}. As illustrated in Figure 2, the progressive decline in lung function (FEV1) begins with a long asymptomatic period during which ongoing functional loss occurs without associated symptoms. Although somewhat variable, the onset of symptoms signifies a substantial loss of function associated with FEV1 of approximately 50% of the predicted normal. One of major challenges in the diagnosis and management of patients with COPD is that individuals typically present at the time of symptom onset when loss of lung function is advanced and interventions such as smoking cessation offer limited clinical improvement. Thus, the ability to identify the subset of smokers who will go on to develop accelerated airflow limitation and chronic respiratory failure while still asymptomatic may result in a profound impact on rate of disease progression and a potential therapeutic window early in the disease process where intervention is more efficacious.

The reason why only a small fraction of tobacco users develop clinically evident COPD remains a fundamental question. A reasonable conclusion is that there are certain subpopulations of tobacco users who have unknown predisposing factors resulting in increased disease susceptibility. Efforts to identify the subpopulations of tobacco users who develop COPD and their associated risk factors remain an active area of COPD research. It is likely that

both environmental and genetic factors have a role in the development of COPD. Environmental factors that have been identified in COPD include smoking, air pollution, tobacco use, and viral infections¹⁶.

Evidence for a genetic link associated with the development of COPD is supported by both twin studies¹⁷ and familial clustering¹⁸. Investigative methods including linkage analysis¹⁸, whole genome screens¹⁹ and candidate gene approaches¹⁹ have been utilized to further identify potential genetic links. From these approaches, a number of candidate genes involved in immune response, cytokine regulation, metabolic and enzymatic processes have been identified. These include alpha-1 antitrypsin, neutrophil elastase, matrix metalloproteases, interleukin-8, tumor necrosis factor- α (TNF- α), transforming growth factor- β (TGF- β), immunoglobulins, vitamin D-binding protein, blood group antigens, and human leukocyte antigen (HLA) status^{19,20}. These data support a role for genetic susceptibility to COPD; however, there likely are multiple genes operating in concert with environmental factors in a synergistic fashion leading to the diverse and complex COPD phenotype.

Although tobacco exposure is the major cause of COPD, smoking cessation does not appear to resolve ongoing airway inflammation once the disease is established⁵. This finding, in combination with the fact that most smokers do not develop COPD, supports the hypothesis that a distinct "immunologic phenotype" exists in smokers who develop COPD. The mechanisms by which the chronic inflammatory response is maintained is not known. However, a distinct "signature" of inflammatory cells, cytokine mediators, and proteases specific to COPD have been identified²¹⁻²³.

The pathologic hallmarks of COPD are persistent airways inflammation, mucus hypersecretion, and lung parenchyma destruction. Studies examining airway histology, bronchoalveolar lavage, and sputum of smokers with COPD have demonstrated increases in macrophages, T-lymphocytes, and neutrophils relative to smokers without COPD, nonsmokers, or asthmatics^{22,24-29}. Clinically, the quantitative presence of airway neutrophils³⁰, macrophages³⁰, and T-lymphocytes²⁰ have been correlated with disease severity, suggesting a role for these immunoregulatory cells in the progression of disease. In particular, the presence of increased CD8+ T-lymphocytes in the airways of COPD patients is a consistent finding^{28,31-33}. The role played by CD8+ T-lymphocytes in underlying disease pathogenesis remains speculative. Current hypotheses include; (1) enhanced apoptosis of alveolar epithelial cells leading to parenchymal destruction³¹, (2) persistent recruitment to the lung parenchyma as a result of recurrent or chronic viral infection causing TNF- α mediated alveolar epithelial cell destruction³⁴, or (3) an autoimmune phenomenon³⁵.

In addition to the abnormal inflammatory response in the airways and parenchyma, there is increasing evidence for systemic (non-pulmonary) inflammation including oxidative stress³⁶, neutrophil activation³⁷⁻³⁹ and apoptosis⁴⁰, and increased proinflammatory cytokines⁴¹⁻⁴³. Several studies have compared peripheral blood T-lymphocytes in smokers versus nonsmokers (Table 1)⁴⁴⁻⁴⁶.

Table 1: Summary of previously performed studies of peripheral blood mononuclear cells in COPD demonstrating role for circulating immunoregulatory cells in the disease process

Study	No	Group	Result
Miller <i>et al.</i>	95	60 smokers vs. 35 nonsmokers	<ul style="list-style-type: none"> • No difference between normals and mild smokers • CD8+, CD4+/CD8+ ratio in heavy smokers vs. normals • reversible with smoking cessation
Ekberg-Jansson <i>et al.</i>	92	58 smokers vs. 34 nonsmokers	<ul style="list-style-type: none"> • T-cell activation marker CD3+ 28+ (p=0.03) smokers vs nonsmokers
De Jong <i>et al.</i>	66	42 current or former smokers with COPD 24 non-COPD smokers or nonsmokers	<ul style="list-style-type: none"> • No difference in T-lymphocyte subsets • CD8+ nonsmoking COPD vs. nonsmoking control (p<0.05) • CD4+/CD8+ ratio associated with FEV1 in nonsmoking COPD (p=0.01)
Kim <i>et al.</i>	60	20 smokers with COPD 20 smokers without COPD 20 healthy nonsmokers	<ul style="list-style-type: none"> • CD8+ (p<0.05) and CD4+/CD8+ ratio (p<0.05) in smokers with COPD and normal DLCO/VA vs. non-COPD smokers
Hodge <i>et al.</i>	34	17 former smokers with COPD 1 current smoker with COPD 16 healthy controls	<ul style="list-style-type: none"> • TNF-α by T-cells (p<0.05) and TNF-α and TGF-β by monocytes (p<0.05 respectively) in COPD vs. controls • annexin V (p=0.006) and Fas expression (p=0.013) in COPD vs. controls
Takabatake <i>et al.</i>	69	34 individuals with COPD 35 healthy controls	<ul style="list-style-type: none"> • TNF-α (p<0.0001) and TNF-receptors (TNF-R55, TNF-R75) (p<0.0001 respectively) in COPD vs. controls

Miller *et al.*⁴⁴ analyzed peripheral blood lymphocyte populations in 60 smokers and 35 nonsmokers. They found that although there was no difference in the total number of T-lymphocytes or CD4+/CD8+ ratios in mild smokers compared to normal individuals, an increase in CD8+ lymphocytes and decreased CD4+/CD8+ ratio was observed in heavy smokers. Interestingly, these changes were reversible with smoking cessation. Ekberg-Jansson *et al.* found that peripheral blood T-cell activating markers were higher in 60 year old male smokers than in age-matched nonsmokers⁴⁶. These studies support the hypothesis that tobacco exposure causes alterations in circulating T-lymphocytes and that these changes may be reversible with smoking cessation⁴⁴. Following these studies, two other groups have investigated the role of smoking and presence or absence of COPD on circulating immunoregulatory T-lymphocytes (Table 1). De Jong *et al.* studied lymphocyte subsets in the

peripheral blood of 42 individuals with COPD both former and active smokers and 24 mixed smoking and nonsmoking subjects without COPD. Results showed no difference in the number of T-lymphocyte subsets when comparing the entire groups or only smoking subjects. However, the percentage of CD8+ T-lymphocytes was significantly higher in the nonsmoking COPD subjects compared to the nonsmoking healthy controls. Additionally, results showed that in the nonsmoking COPD patients, a higher CD4+/CD8+ ratio correlated with improved lung function, again suggesting that the balance of T-lymphocyte subsets may play a role in disease pathogenesis⁴⁷. Kim *et al.* investigated the relationship between T-lymphocyte subsets and pulmonary function in three groups: smokers with COPD, smokers without COPD, and healthy nonsmokers. The authors found that the proportion of CD8+ T-cells and the ratio of CD4+/CD8+ T-cells correlated with physiologic measures of gas exchange⁴⁸. Recently, Hodge *et al.* analyzed peripheral blood T-cells from 18 patients with COPD and found increased T-cell apoptosis and Fas expression compared to normal controls. In addition, intracellular production of TNF- α and TGF- β by T-cells and monocytes in COPD patients was significantly increased over controls⁴⁹. These findings are consistent with the report of Takabatake *et al.* demonstrating increased circulating TNF- α in the peripheral blood of patients with COPD⁵⁰. Additionally, De Godoy *et al.* demonstrated significantly higher levels of TNF- α production in circulating peripheral blood monocytes in a subset of "weight-losing" COPD patients compared to "weight-stable" COPD patients suggesting a possible pathogenetic relationship between accelerated metabolism and TNF- α production⁴³. Thus, there exists strong evidence not only for a quantitative change in T-lymphocyte populations in COPD but also a functional alteration in the circulating immunoregulatory cells (T-lymphocytes and monocytes) that may result in either disease progression, phenotypic variation, or lack of resolution.

Questions remain as to whether the development of COPD in a susceptible smoker is the result of an abnormal inflammatory response to a normal stimulus (i.e. tobacco smoke) or conversely, a normal inflammatory response to an abnormal stimulus (i.e. autoantigen). Arguments for the latter are supported by the ongoing communication and trafficking of inflammatory cells between the pulmonary and systemic circulation. Alternatively, or additionally, there is evidence that variation in T-lymphocyte subpopulations is genetically influenced^{52,53}. Thus, genetic alterations in the host inflammatory response may result the "immunologic phenotype" observed in tobacco related COPD patients.

Chronic obstructive pulmonary disease is a growing worldwide epidemic. Despite the increasing individual and societal costs related to COPD, funding and support for COPD research has been relatively neglected. Thus, the understanding of the pathogenetic and cellular

mechanisms involved in the development of COPD remains limited, and therefore, so do therapeutic approaches for preventing and treating the disease.

Summary of the Invention

5 The present invention provides a method to diagnose chronic obstructive pulmonary disease (COPD) or a predisposition to develop COPD, comprising, detecting the level of expression of at least one gene in a sample of peripheral blood cells isolated from a patient to be tested, wherein the gene is chosen from the genes represented by SEQ ID NO:1-323, and wherein the level of expression of each of the genes in any one or more of Tables 2-5 is associated with COPD as measured by either upregulation or downregulation of gene expression in peripheral blood cells from patients with COPD as compared to the level of expression of the genes in peripheral blood cells from normal controls; and comparing the level of expression of the gene from the patient sample to the level of expression of the gene in normal control peripheral blood cells, wherein detection of regulation of the expression of the gene in the patient sample in the direction associated with COPD indicated in Table 2, 3, 4 and/or 5 indicates a diagnosis of COPD in the patient. In some embodiments, the detecting comprises detecting expression of at least 5 genes chosen from the genes represented by SEQ ID NO:1-323. In some embodiments, the detecting comprises detecting expression of at least 5 genes chosen from the genes represented by SEQ ID NO:1-323. In some embodiments, the detecting comprises detecting expression of at least 10 genes chosen from the genes represented by SEQ ID NO:1-323. In some embodiments, the detecting comprises detecting expression of at least 15 genes chosen from the genes represented by SEQ ID NO:1-323. In some embodiments, the detecting comprises detecting expression of at least 20 genes chosen from the genes represented by SEQ ID NO:1-323. In some embodiments, the detecting comprises detecting expression of at least 25 genes chosen from the genes represented by SEQ ID NO:1-323. In some embodiments, the detecting comprises detecting expression of at least 50 genes chosen from the genes represented by SEQ ID NO:1-323. In some embodiments, the detecting comprises detecting expression of at least 75 genes chosen from the genes represented by SEQ ID NO:1-323. In some embodiments, the detecting comprises detecting expression of at least 100 genes chosen from the genes represented by SEQ ID NO:1-323. In some embodiments, the detecting comprises detecting expression of at least 125 genes chosen from the genes represented by SEQ ID NO:1-323. In some embodiments, the detecting comprises detecting expression of at least 150 genes chosen from the genes represented by SEQ ID NO:1-323. In some embodiments, the detecting comprises detecting expression of at least 175 genes chosen

from the genes represented by SEQ ID NO:1-323. In some embodiments, the detecting comprises detecting expression of at least 200 genes chosen from the genes represented by SEQ ID NO:1-323. In some embodiments, the detecting comprises detecting expression of at least 225 genes chosen from the genes represented by SEQ ID NO:1-323. In some embodiments, the detecting comprises detecting expression of all of the genes represented by SEQ ID NO:1-323.

In some embodiments, expression of the gene is detected by measuring amounts of transcripts of the gene in the patient peripheral blood cells. In other embodiments, expression of the gene is detected by detecting hybridization of at least a portion of the gene or a transcript thereof to a nucleic acid molecule comprising a portion of the gene or a transcript thereof in a nucleic acid array. In some embodiments, expression of the gene is detected by detecting the production of a protein encoded by the gene. In further embodiments, the level of expression of the gene in the peripheral blood cells of a normal control has been predetermined.

The present invention also provides a method to monitor the treatment of a patient with COPD, comprising detecting the level of expression of at least one gene in a sample of peripheral blood cells isolated from a patient undergoing treatment for COPD, wherein the gene is chosen from the genes represented by any one of SEQ ID NO:1-324, and wherein the level of expression of each of the genes represented by any one of SEQ ID NO:1-324 is associated with COPD as measured by either upregulation or downregulation of gene expression in peripheral blood cells from patients with COPD as compared to the level of expression of the genes in peripheral blood cells from normal controls; and comparing the level of expression of the gene from the patient sample to the level of expression of the gene in a prior sample of peripheral blood cells from the patient, wherein detection of a change in the level of expression of the gene, as compared to the level of expression in the prior sample, toward the level of the expression of the gene in a normal control sample, indicates that the treatment for COPD is producing a beneficial result.

In some embodiments, detection of a change in the level of expression of the gene, as compared to the level of expression in the prior sample, away from the level of the expression of the gene in a normal control sample, indicates a progression of the COPD. In other embodiments, the detection of no significant change in the level of expression of the gene, as compared to the level of expression in the prior sample, indicates no significant change in the progression or treatment of the COPD in the patient.

The invention also provides a plurality of polynucleotides for the detection of the expression of genes that are indicative of COPD in a patient or a preclinical disposition therefore, wherein the plurality of polynucleotides consists of at least two polynucleotides,

wherein each polynucleotide is at least 5 nucleotides in length, and wherein each polynucleotide is complementary to an RNA transcript, or nucleotide derived therefrom, of a gene that is regulated differently in individuals with COPD as compared to individuals that do not have COPD.

5 In some embodiments, each polynucleotide is complementary to an RNA transcript, or a polynucleotide derived therefrom, of a gene represented by any one of SEQ ID NO:1-324. In some embodiments, the plurality of polynucleotides comprises polynucleotides that are complementary to an RNA transcript, or a nucleotide derived therefrom, of at least two genes represented by any one of SEQ ID NO:1-324. In some embodiments, the plurality of
10 polynucleotides comprises polynucleotides that are complementary to an RNA transcript, or a nucleotide derived therefrom, of at least five genes represented by any one of SEQ ID NO:1-324. In some embodiments, the plurality of polynucleotides comprises polynucleotides that are complementary to an RNA transcript, or a nucleotide derived therefrom, of at least 10 genes represented by any one of SEQ ID NO:1-324. In some embodiments, the plurality of
15 polynucleotides comprises polynucleotides that are complementary to an RNA transcript, or a nucleotide derived therefrom, of at least 50 genes represented by any one of SEQ ID NO:1-324. In some embodiments, the plurality of polynucleotides comprises polynucleotides that are complementary to an RNA transcript, or a nucleotide derived therefrom, of at least 100 genes represented by any one of SEQ ID NO:1-229, and 320-324. In some embodiments, the plurality of
20 polynucleotides comprises polynucleotides that are complementary to an RNA transcript, or a nucleotide derived therefrom, of at least 150 genes represented by any one of SEQ ID NO:1-229, and 320-324. In some embodiments, the plurality of polynucleotides comprises polynucleotides that are complementary to an RNA transcript, or a nucleotide derived therefrom, of at least 200 genes represented by any one of SEQ ID NO:1-229, and 320-324. In
25 some embodiments, the polynucleotides are immobilized on a substrate. In some embodiments, the polynucleotides are hybridizable array elements in a microarray. In some embodiments, the polynucleotides are conjugated to detectable markers.

The invention further provides a method to diagnose chronic obstructive pulmonary disease (COPD) in a patient, comprising, detecting the level of expression of at least one gene
30 in a sample of peripheral blood cells isolated from a patient to be tested, wherein the gene is chosen from a group of genes, each of which has been previously identified to be upregulated or downregulated in the peripheral blood cells of patients who have been diagnosed with COPD, as compared to the level of expression of the gene in normal control peripheral blood cells; and comparing the level of expression of the gene from the patient sample to the level of expression

of the gene in normal control peripheral blood cells, wherein detection of regulation of the expression of the gene in the patient sample in the direction associated with COPD as indicated by the previous identification, indicates a diagnosis of COPD in the patient.

5 The invention further provides a method to identify a compound with the potential to treat or prevent chronic obstructive pulmonary disease (COPD), comprising contacting a test compound with a cell that expresses a gene selected from any one or more of the genes represented by any one of SEQ ID NO:1-324; identifying compounds that increase the expression or activity of genes represented by any one of SEQ ID NO:1-324 or the proteins encoded thereby that are downregulated in peripheral blood cells of patients with COPD as compared to peripheral blood cells of normal controls, or that decrease the expression or activity of genes represented by any one of SEQ ID NO:1-324 or the proteins encoded thereby that are upregulated in peripheral blood cells of patients with COPD as compared to peripheral blood cells of normal controls.

15 In some embodiments, the invention provides a method to treat a patient with COPD, comprising administering to the patient a therapeutic composition comprising a compound identified by the method above.

20 In some embodiments, detection of a change in the level of expression of at least one gene in methods of the invention comprises detecting the presence of a protein. In other embodiments, the method further comprises detecting the presence of the protein using a reagent that specifically binds to the protein. In some embodiments, the reagent is selected from the group consisting of an antibody, an antibody derivative, and an antibody fragment.

25 The invention also provides a plurality of reagents for the detection of the expression of genes that are indicative of COPD in a patient or a preclinical disposition therefore; wherein the plurality of reagents consists of at least two reagents that each of which specifically bind to a protein, wherein each protein is at least 15 amino acids in length, and wherein each protein is encoded a gene that is regulated differently in individuals with COPD as compared to individuals that do not have COPD.

30 In some embodiments, each protein is encoded a gene represented by any one of SEQ ID NO:1-324. In some embodiments, at least two proteins are encoded a gene represented by any one of SEQ ID NO:1-324. In some embodiments, at least five proteins are encoded a gene represented by any one of SEQ ID NO:1-324. In some embodiments, at least 10 proteins are encoded a gene represented by any one of SEQ ID NO:1-324.. In some embodiments, at least 50 proteins are encoded a gene represented by any one of SEQ ID NO:1-324. In some embodiments, at least 100 proteins are encoded a gene represented by any one of SEQ ID

NO:1-229, and 320-324. In some embodiments, at least 150 proteins are encoded a gene represented by any one of SEQ ID NO:1-229, and 320-324. In some embodiments, at least 200 proteins are encoded a gene represented by any one of SEQ ID NO:1-229, and 320-324. In some embodiments, the reagents are immobilized on a substrate. In some embodiments, the reagents are selected from the group consisting of an antibody, an antibody derivative, and an antibody fragment, and wherein each of said reagents are elements in a microarray. In some embodiments, reagents are conjugated to detectable markers.

Brief Description of the Figures of the Invention

10 Fig. 1 shows the relationship of tobacco exposure to chronic obstructive pulmonary disease (COPD). Only approximately 10-20% of all tobacco exposed individuals develop COPD. However, tobacco exposure accounts for over 90% of COPD in developed countries. *Other causes of COPD include alpha-one antitrypsin deficiency, cystic fibrosis, environmental exposures, etc.

15 Fig. 2 shows examples of normal and accelerated loss of lung function over time. S1 and S2 denote two individuals with tobacco exposure. S1 has accelerated loss of lung function with onset of symptoms at young age resulting in early diagnosis of COPD. S2 also has greater than expected loss of lung function but remains asymptomatic over lifetime without clinical diagnosis of COPD. NS is normal nonsmoker with expected rate of decline in lung function over time.

20 Fig. 3A and B are dendrograms of COPD and Normal samples, clustered using centered correlation and average linkage, and the tile plot of differentially expressed genes. Fig. 3A shows unsupervised clustering based on 15022 genes. Fig. 3B shows supervised clustering based on 240 genes. Fig. 3C show a tile plot of the 240 genes which preliminarily discriminate between COPD and normal PBMC samples. Darker areas represent high expression and lighter areas represent relatively low expression.

Fig. 4 is a schematic drawing showing the study methods used in Example 2.

Fig. 5 is a schematic drawing showing the validation protocol for differentially expressed genes.

30

Description of the Invention

The present invention generally relates to the identification of a large number of genes that are regulated differentially in individuals with chronic pulmonary obstructive disease (COPD) as compared to individuals that do not have this disease, and particularly, to the

identification of *how* these genes are regulated during disease. In addition, this invention generally relates to diagnostic and prognostic assays and kits for COPD, as well as the identification of targets for therapeutic prevention and intervention strategies. According to the present invention, the terms "chronic pulmonary obstructive disease", its acronym "COPD", and "emphysema" can be used interchangeably to describe the same condition.

More specifically, the present inventors have studied gene expression profiles in circulating peripheral blood mononuclear cells (PBMC) as surrogate markers for COPD. Given the systemic inflammatory changes associated with COPD and the knowledge that circulating lymphocytes traffic between the lungs and the peripheral circulation, the present inventors hypothesized that clinically relevant changes in gene expression could be observed in PBMC of COPD patients. The use of gene expression microarray technology is a powerful high-throughput method to investigate relative differences in gene expression between tobacco exposed individuals with and without COPD as well as to identify novel candidate genes and/or candidate therapeutic targets for future study and intervention. PBMC expression profiling is a powerful tool to identify prognostic indicators, disease subclassifications, candidate genes, and response to therapy in a diverse array of disease states⁶⁻¹². However, the global analysis of gene expression utilizing minimally invasive methods in COPD has never been conducted. The present inventors provide data herein showing that there is a distinct gene expression profile among individuals with COPD. The identification of distinct expression profiles between tobacco exposed individuals both with and without COPD is useful for the analysis of disease pathogenesis and immunoregulatory mechanisms, and are further useful as biomarkers that can identify COPD susceptible individuals while in a pre-clinical state.

The use of peripheral blood mononuclear cells (PBMC) as surrogate markers of other diseases has been previously described in the medical literature. PBMC expression analyses have been used to identify prognostic indicators, disease subclassifications, candidate genes, and response to therapy in disease states including acute leukemia^{6,7}, sickle cell disease⁹, connective tissue disease^{8,10}, multiple sclerosis¹¹, and colorectal cancer¹². Recently, the present inventors and colleagues published a report using PBMC expression profiling to identify differentially expressed genes between idiopathic pulmonary arterial hypertension and secondary pulmonary hypertension. These studies showed that one can prospectively discriminate between patients with pulmonary hypertension and normal controls based solely on expression profiles, thus supporting the concept of an "immunologic signature" associated with a certain disease process¹³.

However, as discussed above, global analysis of peripheral blood gene expression in COPD had never been conducted prior to the present invention. Moreover, it is not necessarily intuitive that circulating blood cells will carry information related to COPD. T lymphocytes, which recognize antigens presented by PBMCs, respond to specific antigens, but not with a disease-specific response to a single antigen. Therefore, while screening PBMCs might be useful to distinguish between normal patients and those with one type of disease, the application of such technology to distinguish between normal patients and those with an entirely different disease is not intuitively operable, nor has such an approach been validated for COPD until the present invention.

Current evidence supports a role for both genetic and environmental factors in COPD disease pathogenesis. However, past focus has been limited to a few genes and/or inflammatory mediators of known interest. Although these approaches are valid, without being bound by theory, the present inventors believe that the study of one or two genes in a population of 5000 individuals may be less informative about the disease process than the study of 30,000 genes in fewer affected individuals. The use of gene expression microarrays to study tobacco exposed individuals both with and without COPD enables the comparison of thousands of expression transcripts between groups and has resulted in the discovery of novel genes of interest, new diagnostic tools, disease subclassifications, and new candidate therapeutic targets. The present inventors' analysis methods have the advantage of generating new hypotheses and investigative pathways based on the study of fewer individuals through minimally invasive methods. The ability to conduct research in human subjects and not just animal models is further likely to enhance the discovery and understanding of this common and underappreciated disease within the target population. The large amounts of data that can be generated from a microarray gene expression study enable the inventors to capture many simultaneous processes and convert these findings into meaningful, quantitative, and reproducible data.

In addition to the discovery of biomarkers that can be used individually or in any combination in assays and kits for the diagnosis of, prognosis of, or other evaluation or study of COPD, the present inventors have uncovered a number of genes not previously recognized to play a role in the disease process of COPD, which can now be studied in more detail and/or be used as targets for the discovery of other modulators of disease or therapeutic agents.

The present inventors have identified multiple genes, the expression of which is regulated differentially in peripheral blood cells (PBC; also referred to herein as peripheral blood mononuclear cells, or PBMC) of patients with COPD as compared to subjects without

COPD. One set of genes identified as being differentially expressed in patients with COPD versus normal controls are shown in Table 2.

Table 2. Genes Identified As Being Differentially Expressed In Patients With COPD Versus Normal Controls

P-value	Normal	COPD	Fold Change	Probe set ID	Name	Symbol
6.9130×10^{-4}	9.2	112.7	0.082	217572_at	SEQ ID NO:320	
3.1620×10^{-3}	16.2	67.7	0.239	202718_at	insulin-like growth factor binding protein 2, 36kDa; SEQ ID NO:1	IGFBP2
2.9800×10^{-5}	112.5	437.9	0.257	208524_at	G protein-coupled receptor 15; SEQ ID NO:2	GPR15
4.2400×10^{-4}	61.1	211.5	0.289	204555_s_at	protein phosphatase 1, regulatory subunit 3D; SEQ ID NO:3	PPP1R3D
8.0110×10^{-3}	125.9	429.9	0.293	211791_s_at	potassium voltage-gated channel, shaker-related subfamily, beta member 2; SEQ ID NO:4	KCNAB2
9.8156×10^{-3}	10.2	34.2	0.298	220609_at	CDNA FLJ42140 fis, clone TESTI2041734; SEQ ID NO:5	
9.9543×10^{-3}	23.8	78.8	0.302	216139_s_at	mitogen-activated protein kinase 8 interacting protein 3; SEQ ID NO:6	MAPK8IP3
7.4840×10^{-3}	13.3	40.3	0.33	219668_at	ganglioside-induced differentiation-associated protein 1-like 1; SEQ ID NO:7	GDAP1L1
6.7123×10^{-3}	78.4	218.2	0.359	201796_s_at	valyl-tRNA synthetase 2; SEQ ID NO:8	VAR2
8.3174×10^{-3}	235.2	649.6	0.362	202072_at	heterogeneous nuclear ribonucleoprotein L; SEQ ID NO:9	HNRPL
4.4919×10^{-3}	45.9	125.1	0.367	208499_s_at	DnaJ (Hsp40) homolog, subfamily C, member 3; SEQ ID NO:10	DNAJC3
2.9942×10^{-3}	147.5	381.6	0.387	204713_s_at	coagulation factor V (proaccelerin, labile factor); SEQ ID NO:11	F5
6.9640×10^{-4}	53.1	135.6	0.392	219978_s_at	nucleolar and spindle associated protein 1; SEQ ID NO:12	NUSAP1
2.1734×10^{-3}	135	343.3	0.393	212451_at	KIAA0256 gene product; SEQ ID NO:12	KIAA0256
9.2912×10^{-3}	27.6	67.2	0.411	220400_at	Cohen syndrome 1; SEQ ID NO:14	COH1
4.9816×10^{-3}	35.2	79.7	0.442	217557_s_at	carboxypeptidase M; SEQ ID NO:15	CPM
3.2933×10^{-3}	156.5	351.3	0.445	208859_s_at	alpha thalassemia/mental retardation syndrome X-linked (RAD54 homolog, S. cerevisiae); SEQ ID NO:16	ATRX
5.5411×10^{-3}	331.1	739.3	0.448	200702_s_at	DEAD (Asp-Glu-Ala-Asp) box polypeptide 24; SEQ ID NO:17	DDX24
4.1850×10^{-3}	313.5	697.9	0.449	201867_s_at	transducin (beta)-like 1X-linked; SEQ ID NO:18	TBL1X
7.4144×10^{-3}	306.9	667.9	0.459	213710_s_at	calmodulin 1 (phosphorylase kinase, delta); SEQ ID NO:19	CALM1
8.6879×10^{-3}	463.1	1003.3	0.462	202379_s_at	natural killer-tumor recognition sequence; SEQ ID NO:20	NKTR
3.4175×10^{-3}	34.3	72.4	0.474	205768_s_at	solute carrier family 27 (fatty acid transporter), member 2; SEQ ID NO:21	SLC27A2
1.0626×10^{-3}	99.8	208.7	0.478	221705_s_at	hypothetical protein FLJ21168; SEQ ID NO:22	FLJ21168
3.5276×10^{-3}	1144	2376.7	0.481	212794_s_at	KIAA1033 protein; SEQ ID NO:23	KIAA1033

P-value	Normal	COP D	Fold Change	Probe set ID	Name	Symbol
9.6802×10^{-3}	135.8	278.1	0.488	218859_s_at	chromosome 20 open reading frame 6; SEQ ID NO:24	C20orf6
9.8882×10^{-3}	93.9	191.7	0.49	219983_at	HRAS-like suppressor ; SEQ ID NO:25	HRASLS
3.2903×10^{-3}	234.8	477.1	0.492	211635_x_at	immunoglobulin heavy constant gamma 1 (G1m marker); SEQ ID NO:26	IGHG1
4.6102×10^{-3}	119	241.8	0.492	203628_at	insulin-like growth factor 1 receptor; SEQ ID NO:27	IGF1R
6.0046×10^{-3}	1165.8	2363.9	0.493	205483_s_at	interferon, alpha-inducible protein (clone IFI-15K); SEQ ID NO:28	G1P2
5.3238×10^{-3}	225.6	451.8	0.499	221524_s_at	Ras-related GTP binding D; SEQ ID NO:29	RRAGD
2.3757×10^{-3}	273.2	543.3	0.503	209088_s_at	ubiquitin 1; SEQ ID NO:30	UBN1
9.4040×10^{-4}	425.6	823.8	0.517	212028_at	RNA binding motif protein 25; SEQ ID NO:31	RBM25
7.2423×10^{-3}	214.2	412.7	0.519	200616_s_at	KIAA0152 gene product; SEQ ID NO:32	KIAA0152
9.5507×10^{-3}	542.6	1008.1	0.538	216915_s_at	protein tyrosine phosphatase, non-receptor type 12; SEQ ID NO:33	PTPN12
7.3134×10^{-3}	498.8	923.3	0.54	208993_s_at	peptidyl-prolyl isomerase G (cyclophilin G); SEQ ID NO:34	PPIG
4.1156×10^{-3}	239.9	436.5	0.55	219865_at	HSPC157 protein; SEQ ID NO:35	HSPC157
8.7175×10^{-3}	175.5	319	0.55	202476_s_at	tubulin, gamma complex associated protein 2; SEQ ID NO:36	TUBGCP2
7.2736×10^{-3}	231.8	420.9	0.551	202503_s_at	KIAA0101 gene product; SEQ ID NO:37	KIAA0101
2.8803×10^{-3}	73.5	133.1	0.552	216541_x_at	immunoglobulin heavy constant gamma 1 (G1m marker); SEQ ID NO:323	IGHG1
4.2567×10^{-3}	115.5	207.9	0.556	205585_at	ets variant gene 6 (TEL oncogene) ; SEQ ID NO:38	ETV6
1.2060×10^{-3}	88.9	156.9	0.567	215300_s_at	flavin containing monooxygenase 5; SEQ ID NO:39	FMO5
6.1886×10^{-3}	1552.9	2740.1	0.567	214352_s_at	v-Ki-ras2 Kirsten rat sarcoma 2 viral oncogene homolog; SEQ ID NO:40	KRAS2
8.7540×10^{-4}	633.8	1116.3	0.568	35201_at	heterogeneous nuclear ribonucleoprotein L; SEQ ID NO:9	HNRPL
7.8950×10^{-3}	474.6	832.2	0.57	213015_at	bobby sox homolog (Drosophila); SEQ ID NO:41	BBX
6.5953×10^{-3}	110.5	193.2	0.572	216563_at	ankyrin repeat domain 12; SEQ ID NO:42	ANKRD12
2.4885×10^{-3}	336.9	584	0.577	204211_x_at	protein kinase, interferon-inducible double stranded RNA dependent ; SEQ ID NO:43	PRKR
9.8671×10^{-3}	264.3	458.1	0.577	202082_s_at	SEC14-like 1 (S. cerevisiae); SEQ ID NO:44	SEC14L1
4.4610×10^{-3}	270.1	464.9	0.581	215357_s_at	polymerase (DNA-directed), delta interacting protein 3; SEQ ID NO:45	POLDIP3
7.0680×10^{-3}	84.1	144.6	0.582	220565_at	G protein-coupled receptor 2; SEQ ID NO:46	GPR2
6.9198×10^{-3}	540.2	921.9	0.586	218302_at	presenilin enhancer 2; SEQ ID NO:47	PEN2

P-value	Normal	COPD	Fold Change	Probe set ID	Name	Symbol
6.0460×10^{-3}	212.7	360.4	0.59	222045_s_at	chromosome 20 open reading frame 67; SEQ ID NO:48	C20orf67
5.3222×10^{-3}	794.4	1337.7	0.594	201730_s_at	translocated promoter region (to activated MET oncogene); SEQ ID NO:49	TPR
4.4620×10^{-4}	143.8	241.8	0.595	212084_at	testis expressed gene 261; SEQ ID NO:50	TEX261
7.2622×10^{-3}	44.9	73.8	0.608	210704_at	fasciculation and elongation protein zeta 2 (zygin II); SEQ ID NO:51	FEZ2
5.3344×10^{-3}	178.5	290.6	0.614	212417_at	secretory carrier membrane protein 1; SEQ ID NO:52	SCAMP1
3.4060×10^{-3}	1143.1	1858.9	0.615	204214_s_at	RAB32, member RAS oncogene family; SEQ ID NO:53	RAB32
8.1933×10^{-3}	858.7	1396.4	0.615	220052_s_at	TERF1 (TRF1)-interacting nuclear factor 2; SEQ ID NO:54	TINF2
6.0307×10^{-3}	802.6	1303.4	0.616	215780_s_at	SET translocation (myeloid leukemia-associated); SEQ ID NO:55	SET
6.1735×10^{-3}	1181	1913	0.617	213982_s_at	expressed in hematopoietic cells, heart, liver; SEQ ID NO:56	HHL
2.6694×10^{-3}	1762.7	2849.3	0.619	213729_at	formin binding protein 3; SEQ ID NO:57	FNBP3
7.7845×10^{-3}	382.6	618.3	0.619	201683_x_at	chromosome 14 open reading frame 92; SEQ ID NO:58	C14orf92
5.5533×10^{-3}	145.1	234	0.62	210241_s_at	TP53 activated protein 1; SEQ ID NO:59	TP53AP1
5.7080×10^{-3}	67.6	109.1	0.62	218009_s_at	protein regulator of cytokinesis 1; SEQ ID NO:60	PRC1
6.5143×10^{-3}	345.2	556.6	0.62	218135_at	PTX1 protein; SEQ ID NO:61	PTX1
6.3304×10^{-3}	1100.5	1772	0.621	215075_s_at	growth factor receptor-bound protein 2 ; SEQ ID NO:62	GRB2
4.4180×10^{-3}	399.5	641.8	0.622	211622_s_at	ADP-ribosylation factor 3; SEQ ID NO:63	ARF3
3.6099×10^{-3}	1410.5	2256.7	0.625	206631_at	prostaglandin E receptor 2 (subtype EP2), 53kDa; SEQ ID NO:64	PTGER2
1.8761×10^{-3}	37.1	58.9	0.63	211249_at	G protein-coupled receptor 68; SEQ ID NO:65	GPR68
4.1248×10^{-3}	271.6	430.9	0.63	212235_at	plexin D1; SEQ ID NO:66	PLXND1
4.6137×10^{-3}	207.9	327.5	0.635	209113_s_at	high-mobility group 20B; SEQ ID NO:67	HMG20B
6.5358×10^{-3}	265.1	413.6	0.641	203771_s_at	biliverdin reductase A; SEQ ID NO:68	BLVRA
9.7479×10^{-3}	2606.2	4054.6	0.643	221479_s_at	BCL2/adenovirus E1B 19kDa interacting protein 3-like; SEQ ID NO:69	BNIP3L
9.3101×10^{-3}	169	261.2	0.647	209743_s_at	itchy homolog E3 ubiquitin protein ligase (mouse); SEQ ID NO:70	ITCH
8.4743×10^{-3}	83	127.6	0.65	219905_at	erythroblast membrane-associated protein; SEQ ID NO:71	ERMAP
7.3512×10^{-3}	898.2	1377	0.652	213418_at	heat shock 70kDa protein 6 (HSP70B'); SEQ ID NO:72	HSPA6

P-value	Normal	COP	Fold	Probe set	Name	Symbol
	1	D	Change	ID		
9.1424×10^{-3}	145.7	223.2	0.653	209890_at	transmembrane 4 superfamily member 9 ; SEQ ID NO:73	TM4SF9
8.7318×10^{-3}	332.5	508.5	0.654	212003_at	putative MAPK activating protein PM20,PM21; SEQ ID NO:74	DKFZp566C0424
5.9490×10^{-3}	1922.5	2936.8	0.655	200744_s_at	guanine nucleotide binding protein (G protein), beta polypeptide 1; SEQ ID NO:75	GNB1
5.1448×10^{-3}	118.5	180	0.658	205538_at	coronin, actin binding protein, 2A; SEQ ID NO:76	CORO2A
5.2360×10^{-3}	127	192	0.661	209943_at	F-box and leucine-rich repeat protein 4 ; SEQ ID NO:77	FBXL4
6.6322×10^{-3}	2380.1	3598.7	0.661	202858_at	U2(RNU2) small nuclear RNA auxiliary factor 1; SEQ ID NO:78	U2AF1
5.4663×10^{-3}	617.5	926.4	0.667	216591_s_at	succinate dehydrogenase complex, subunit C, integral membrane protein, 15kDa; SEQ ID NO:79	SDHC
3.6738×10^{-3}	14232.3	21276.6	0.669	203535_at	S100 calcium binding protein A9 (calgranulin B); SEQ ID NO:80	S100A9
6.6455×10^{-3}	96	143.1	0.671	212128_s_at	dystroglycan 1 (dystrophin-associated glycoprotein 1); SEQ ID NO:81	DAG1
2.1096×10^{-3}	426	630.3	0.676	207216_at	tumor necrosis factor (ligand) superfamily, member 8; SEQ ID NO:82	TNFSF8
6.1573×10^{-3}	297.2	438.4	0.678	204084_s_at	ceroid-lipofuscinosis, neuronal 5; SEQ ID NO:83	CLN5
2.6596×10^{-3}	92.8	136	0.682	213945_s_at	nucleoporin 210; SEQ ID NO:84	NUP210
7.2394×10^{-3}	849.9	1239.4	0.686	210992_x_at	Fc fragment of IgG, low affinity IIa, receptor for (CD32); SEQ ID NO:85	FCGR2A
6.2429×10^{-3}	342.6	497.1	0.689	219382_at	SERTA domain containing 3; SEQ ID NO:86	SERTAD3
2.5365×10^{-3}	1365.7	1978	0.69	210131_x_at	succinate dehydrogenase complex, subunit C, integral membrane protein, 15kDa; SEQ ID NO:79	SDHC
3.0598×10^{-3}	1794.5	2602.1	0.69	200900_s_at	mannose-6-phosphate receptor (cation dependent); SEQ ID NO:87	M6PR
5.2622×10^{-3}	697.5	1009	0.691	217909_s_at	transcription factor-like 4; SEQ ID NO:88	TCFL4
5.2730×10^{-3}	219	314.8	0.696	205045_at	A kinase (PRKA) anchor protein 10; SEQ ID NO:89	AKAP10
7.7137×10^{-3}	87.1	125.1	0.696	206341_at	interleukin 2 receptor, alpha; SEQ ID NO:90	IL2RA
5.6831×10^{-3}	2063.7	2955.7	0.698	201593_s_at	likely ortholog of mouse immediate early response, erythropoietin 4; SEQ ID NO:91	LEREPO4
5.9180×10^{-4}	7748.4	10972.5	0.706	211750_x_at	tubulin alpha 6; SEQ ID NO:92	TUBA6
7.3069×10^{-3}	422.7	590.7	0.716	205349_at	guanine nucleotide binding protein (G protein), alpha 15 (Gq class) ; SEQ ID NO:93	GNA15
9.9014×10^{-3}	155.4	216.7	0.717	208701_at	amyloid beta (A4) precursor-like protein 2; SEQ ID NO:94	APLP2

P-value	Norma l	COP D	Fold Chang e	Probe set ID	Name	Symbol
4.3947×10^{-3}	13844. 3	19228	0.72	217398_x_ at	glyceraldehyde-3-phosphate dehydrogenase; SEQ ID NO:95	GAPD
6.4329×10^{-3}	12225. 4	16959. 3	0.721	213453_x_ at	glyceraldehyde-3-phosphate dehydrogenase; SEQ ID NO:95	GAPD
9.1361×10^{-3}	1969.5	2732.4	0.721	212420_at	E74-like factor 1 (ets domain transcription factor); SEQ ID NO:96	ELF1
4.5457×10^{-3}	602.7	833.7	0.723	216326_s_ at	histone deacetylase 3; SEQ ID NO:97	HDAC3
9.5438×10^{-3}	532.4	733.2	0.726	202301_s_ at	similar to splicing factor, arginine/serine-rich 4; SEQ ID NO:98	FLJ11021
6.8390×10^{-3}	5247.4	7203.8	0.728	217356_s_ at	phosphoglycerate kinase 1; SEQ ID NO:99	PGK1
4.6545×10^{-3}	2072.5	2835.5	0.731	211025_x_ at	cytochrome c oxidase subunit Vb ; SEQ ID NO:100	COX5B
8.0408×10^{-3}	312.3	426.5	0.732	222216_s_ at	mitochondrial ribosomal protein L17; SEQ ID NO:101	MRPL17
3.3102×10^{-3}	2213.4	2971.7	0.745	205159_at	colony stimulating factor 2 receptor, beta, low- affinity (granulocyte-macrophage); SEQ ID NO:102	CSF2RB
6.7899×10^{-3}	14348. 6	19180. 5	0.748	212581_x_ at	glyceraldehyde-3-phosphate dehydrogenase; SEQ ID NO:95	GAPD
7.2782×10^{-3}	147.6	196.9	0.75	203163_at	katanin p80 (WD repeat containing) subunit B 1; SEQ ID NO:103	KATNB1
3.8572×10^{-3}	10506. 8	13890. 9	0.756	214394_x_ at	eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange protein); SEQ ID NO:104	EEF1D
4.4130×10^{-3}	4386.4	5782.3	0.759	212716_s_ at	eukaryotic translation initiation factor 3 subunit k; SEQ ID NO:105	eIF3k
9.9862×10^{-3}	3424.9	4463.6	0.767	201441_at	cytochrome c oxidase subunit VIb ; SEQ ID NO:106	COX6B
6.4834×10^{-3}	450.2	577.3	0.78	200913_at	protein phosphatase 1G (formerly 2C), magnesium-dependent, gamma isoform ; SEQ ID NO:107	PPM1G
8.6102×10^{-3}	110	85.1	1.293	212719_at	pleckstrin homology domain containing, family E (with leucine rich repeats) member 1; SEQ ID NO:108	PLEKHE1
7.6256×10^{-3}	1639	1257.3	1.304	211971_s_ at	leucine-rich PPR-motif containing; SEQ ID NO:109	LRPPRC
6.1002×10^{-3}	645.1	486.5	1.326	202474_s_ at	host cell factor C1 (VP16-accessory protein) ; SEQ ID NO:110	HCFC1
5.9689×10^{-3}	515.9	388.3	1.329	205323_s_ at	metal-regulatory transcription factor 1 ; SEQ ID NO:111	MTF1
6.8278×10^{-3}	164.7	122.5	1.344	212744_at	Bardet-Biedl syndrome 4; SEQ ID NO:112	BBS4
5.0314×10^{-3}	470.1	348.9	1.347	203566_s_ at	amylo-1, 6-glucosidase, 4-alpha- glucanotransferase (glycogen debranching enzyme, glycogen storage disease type III); SEQ ID NO:113	AGL
7.3403×10^{-3}	456.9	338.6	1.349	213542_at	hypothetical protein DKFZp547K1113; SEQ ID NO:114	DKFZp547 K1113

P-value	Normal	COP D	Fold Change	Probe set ID	Name	Symbol
3.5617×10^{-3}	575.2	421.8	1.364	201908_at	dishevelled, dsh homolog 3 (Drosophila) ; SEQ ID NO:115	DVL3
3.3610×10^{-3}	5998.7	4362	1.375	208997_s_at	uncoupling protein 2 (mitochondrial, proton carrier); SEQ ID NO:116	UCP2
8.2070×10^{-3}	459.7	333.7	1.378	219628_at	p53 target zinc finger protein; SEQ ID NO:117	WIG1
5.4116×10^{-3}	818.7	591.6	1.384	201075_s_at	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily c, member 1; SEQ ID NO:118	SMARCC1
9.6285×10^{-3}	853	614.9	1.387	202761_s_at	spectrin repeat containing, nuclear envelope 2; SEQ ID NO:119	SYNE2
5.8320×10^{-3}	191.9	138.2	1.389	205139_s_at	uronyl-2-sulfotransferase; SEQ ID NO:120	UST
9.9572×10^{-3}	817.2	588	1.39	221428_s_at	likely ortholog of mouse IRA1 protein; SEQ ID NO:121	IRA1
9.3746×10^{-3}	436	312.6	1.395	221744_at	WD-repeat protein; SEQ ID NO:122	HAN11
8.2675×10^{-3}	1111.4	796.1	1.396	209110_s_at	RAB2, member RAS oncogene family-like ; SEQ ID NO:123	RAB2L
5.5122×10^{-3}	1126.1	801.3	1.405	208884_s_at	progesterone induced protein; SEQ ID NO:124	DD5
2.2874×10^{-3}	616.4	435.5	1.415	215714_s_at	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4; SEQ ID NO:125	SMARCA4
7.8899×10^{-3}	381.9	269.8	1.415	208989_s_at	F-box and leucine-rich repeat protein 11; SEQ ID NO:126	FBXL11
7.4646×10^{-3}	451.5	318.9	1.416	216885_s_at	H326; SEQ ID NO:127	H326
5.9775×10^{-3}	1137.7	801.5	1.419	219279_at	dedicator of cytokinesis 10; SEQ ID NO:128	DOCK10
6.0046×10^{-3}	100.3	70.7	1.419	220467_at	hypothetical protein FLJ21272; SEQ ID NO:129	FLJ21272
6.5556×10^{-3}	570.9	398.9	1.431	216071_x_at	trinucleotide repeat containing 11 (THR-associated protein, 230kDa subunit); SEQ ID NO:130	TNRC11
3.0739×10^{-3}	108.1	75.4	1.434	205088_at	chromosome X open reading frame 6 ; SEQ ID NO:131	CXorf6
3.2203×10^{-3}	320.6	223.6	1.434	203183_s_at	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 1; SEQ ID NO:132	SMARCD1
4.7407×10^{-3}	227.1	157.8	1.439	219880_at	hypothetical protein FLJ23053; SEQ ID NO:133	FLJ23053
5.5411×10^{-3}	1210	837	1.446	214315_x_at	calreticulin; SEQ ID NO:134	CALR
4.5034×10^{-3}	119.2	82.3	1.448	200688_at	splicing factor 3b, subunit 3, 130kDa; SEQ ID NO:135	SF3B3
5.6222×10^{-3}	1040	717.8	1.449	206542_s_at	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 2; SEQ ID NO:136	SMARCA2
8.0935×10^{-3}	393.7	271.7	1.449	209998_at	phosphatidylinositol glycan, class O; SEQ ID NO:137	PIGO

P-value	Norma I	COP D	Fold Chang e	Probe set ID	Name	Symbol
6.8661 x 10 ⁻³	1188.6	818.4	1.452	203297_s_at	Jumonji, AT rich interactive domain 2; SEQ ID NO:138	JARID2
7.7196 x 10 ⁻³	720.7	495.2	1.455	212842_x_at	GRIP and coiled-coil domain containing 2; SEQ ID NO:139	GCC2
7.0965 x 10 ⁻³	178.2	120.9	1.474	201929_s_at	plakophilin 4; SEQ ID NO:140	PKP4
3.4954 x 10 ⁻³	185.2	125.6	1.475	204291_at	zinc finger protein 518; SEQ ID NO:141	ZNF518
9.8926 x 10 ⁻³	231.5	156.6	1.478	209084_s_at	replication factor C (activator 1) 1, 145kDa; SEQ ID NO:142	RFC1
3.8876 x 10 ⁻³	1682	1134.8	1.482	208875_s_at	p21 (CDKN1A)-activated kinase 2; SEQ ID NO:143	PAK2
3.6852 x 10 ⁻³	309.6	207.4	1.493	212309_at	cytoplasmic linker associated protein 2; SEQ ID NO:144	CLASP2
4.2610 x 10 ⁻⁴	170.5	114	1.496	214691_x_at	hypothetical protein KIAA1164; SEQ ID NO:145	KIAA1164
3.7728 x 10 ⁻³	622.2	415.3	1.498	203248_at	zinc finger protein 24 (KOX 17); SEQ ID NO:146	ZNF24
1.6424 x 10 ⁻³	93.2	62.1	1.501	207064_s_at	amine oxidase, copper containing 2 (retina-specific); SEQ ID NO:147	AOC2
3.1603 x 10 ⁻³	407.1	271.2	1.501	213018_at	ocular development-associated gene; SEQ ID NO:148	ODAG
7.6665 x 10 ⁻³	155.4	103.5	1.501	221332_at	bone morphogenetic protein 15; SEQ ID NO:149	BMP15
3.3552 x 10 ⁻³	217.1	144.3	1.505	206474_at	PCTAIRE protein kinase 2; SEQ ID NO:150	PCTK2
9.7840 x 10 ⁻⁴	299.4	198.8	1.506	212218_s_at	F-box only protein 9; SEQ ID NO:151	FBXO9
7.3922 x 10 ⁻³	745.6	494.8	1.507	202519_at	Mlx interactor; SEQ ID NO:152	MONDOA
8.5511 x 10 ⁻³	181.7	120.2	1.512	216226_at	TAF4b RNA polymerase II, TATA box binding protein (TBP)-associated factor, 105kDa; SEQ ID NO:153	TAF4B
9.0227 x 10 ⁻³	724.6	479.1	1.512	201307_at	hypothetical protein FLJ10849; SEQ ID NO:154	FLJ10849
7.1990 x 10 ⁻⁴	897.8	593.5	1.513	203480_s_at	HIV-1 induced protein HIN-1; SEQ ID NO:155	HSN1
9.2536 x 10 ⁻³	1590.4	1045.3	1.521	208598_s_at	upstream regulatory element binding protein 1; SEQ ID NO:156	UREB1
6.5580 x 10 ⁻⁴	322.1	211.5	1.523	201439_at	golgi-specific brefeldin A resistance factor 1; SEQ ID NO:157	GBF1
4.4773 x 10 ⁻³	127.6	83.8	1.523	219901_at	FYVE, RhoGEF and PH domain containing 6; SEQ ID NO:158	FGD6
4.9593 x 10 ⁻³	510.5	332.1	1.537	205434_s_at	AP2 associated kinase 1; SEQ ID NO:159	AAK1
6.2381 x 10 ⁻³	73.8	48	1.537	207417_s_at	zinc finger protein 177; SEQ ID NO:160	ZNF177
3.9298 x 10 ⁻³	2645.3	1718	1.54	204197_s_at	runt-related transcription factor 3 ; SEQ ID NO:161	RUNX3
2.0799 x 10 ⁻³	823.2	531.8	1.548	215235_at	spectrin, alpha, non-erythrocytic 1 (alpha-fodrin); SEQ ID NO:162	SPTAN1
2.5208 x 10 ⁻³	390.4	251.5	1.552	219269_at	hypothetical protein FLJ21616; SEQ ID	FLJ21616

P-value	Normal	COPD	Fold Change	Probe set ID	Name	Symbol
					NO:163	
7.8821×10^{-3}	62.8	40.4	1.554	222166_at	chromosome 9 open reading frame 16; SEQ ID NO:164	C9orf16
3.1227×10^{-3}	241.7	154.5	1.564	212822_at	HEG homolog; SEQ ID NO:165	HEG
8.2000×10^{-3}	249.6	158.3	1.577	215260_s_at	transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47); SEQ ID NO:166	TCF3
4.5920×10^{-4}	152.5	96.6	1.579	207394_at	zinc finger protein 137 (clone pHZ-30); SEQ ID NO:167	ZNF137
6.6503×10^{-3}	100.1	63.2	1.584	220198_s_at	eukaryotic translation initiation factor 5A2 ; SEQ ID NO:168	EIF5A2
3.9896×10^{-3}	214.6	134.9	1.591	210739_x_at	solute carrier family 4, sodium bicarbonate cotransporter, member 4; SEQ ID NO:169	SLC4A4
2.5114×10^{-3}	512.7	321.6	1.594	202249_s_at	H326; SEQ ID NO:127	H326
4.4671×10^{-3}	194.8	122.2	1.594	218507_at	hypoxia-inducible protein 2; SEQ ID NO:170	HIG2
2.7104×10^{-3}	84.7	53.1	1.595	210331_at	HECT type E3 ubiquitin ligase; SEQ ID NO:171	NEDL1
1.4838×10^{-3}	164.1	102.7	1.598	210796_x_at	sialic acid binding Ig-like lectin 6; SEQ ID NO:172	SIGLEC6
7.8310×10^{-3}	41.7	26.1	1.598	217997_at	pleckstrin homology-like domain, family A, member 1; SEQ ID NO:173	PHLDA1
4.6176×10^{-3}	307.2	191.8	1.602	217902_s_at	hect domain and RLD 2; SEQ ID NO:174	HERC2
3.2295×10^{-3}	537.8	334.5	1.608	217991_x_at	single stranded DNA binding protein 3; SEQ ID NO:175	SSBP3
1.1893×10^{-3}	440.3	271.1	1.624	212772_s_at	ATP-binding cassette, sub-family A (ABC1), member 2; SEQ ID NO:176	ABCA2
6.5640×10^{-4}	122.4	75.3	1.625	207498_s_at	cytochrome P450, family 2, subfamily D, polypeptide 6; SEQ ID NO:177	CYP2D6
4.8866×10^{-3}	75.9	46.6	1.629	219941_at	transmembrane protein 19; SEQ ID NO:178	TMEM19
5.0786×10^{-3}	54.9	33.7	1.629	222310_at	splicing factor, arginine/serine-rich 15; SEQ ID NO:179	SFRS15
9.1863×10^{-3}	1277.2	783.8	1.629	200601_at	actinin, alpha 4 ; SEQ ID NO:180	ACTN4
5.7037×10^{-3}	181.9	110	1.654	214081_at	plexin domain containing 1; SEQ ID NO:181	PLXDC1
7.8480×10^{-3}	120.9	72.8	1.661	213484_at	Clone 23700 mRNA sequence; SEQ ID NO:182	
3.7009×10^{-3}	130.5	78.5	1.662	210486_at	ankyrin repeat and MYND domain containing 1; SEQ ID NO:183	ANKMY1
9.2644×10^{-3}	28.5	17.1	1.667	211740_at	islet cell autoantigen 1, 69kDa; SEQ ID NO:184	ICA1
9.9970×10^{-3}	413.9	247.9	1.67	206828_at	TXK tyrosine kinase; SEQ ID NO:185	TXK
4.0838×10^{-3}	80.4	48.1	1.672	210290_at	zinc finger protein 174; SEQ ID NO:186	ZNF174
6.2980×10^{-4}	1560	928.3	1.68	210649_s_at	AT rich interactive domain 1A (SWI-like); SEQ ID NO:187	ARID1A
5.5184×10^{-3}	186.7	110.8	1.685	213216_at	KIAA0459 protein; SEQ ID NO:188	KIAA0459

P-value	Normal	COPD	Fold Change	Probe set ID	Name	Symbol
3.3556×10^{-3}	37.7	21.8	1.729	206847_s_at	homeo box A7; SEQ ID NO:189	HOXA7
3.1931×10^{-3}	873.9	503.4	1.736	212553_at	KIAA0460 protein; SEQ ID NO:190	KIAA0460
7.1398×10^{-3}	45	25.9	1.737	214569_at	interferon, alpha 5; SEQ ID NO:191	IFNA5
3.0534×10^{-3}	45.5	26	1.75	216590_at	SEQ ID NO:321	
5.2333×10^{-3}	93	53.1	1.751	220869_at	SEQ ID NO:322	
2.8630×10^{-4}	492.8	280.6	1.756	214659_x_at	chromosome 14 open reading frame 170; SEQ ID NO:192	C14orf170
2.7750×10^{-3}	88.2	50.1	1.76	215135_at	aspartyl aminopeptidase; SEQ ID NO:193	DNPEP
9.2200×10^{-4}	107.9	60.4	1.786	214947_at	Clone 24653 mRNA sequence; SEQ ID NO:194	
2.7649×10^{-3}	80	44.4	1.802	206557_at	hypothetical protein FLJ12985; SEQ ID NO:195	FLJ12985
5.8607×10^{-3}	86.2	47.6	1.811	209816_at	patched homolog (Drosophila); SEQ ID NO:196	PTCH
1.2411×10^{-3}	89.9	49.6	1.812	207950_s_at	ankyrin 3, node of Ranvier (ankyrin G); SEQ ID NO:197	ANK3
3.2337×10^{-3}	167	91.8	1.819	207188_at	cyclin-dependent kinase 3 ; SEQ ID NO:198	CDK3
9.1443×10^{-3}	61.9	33.7	1.837	215295_at	dystrobrevin, beta; SEQ ID NO:199	DTNB
2.9277×10^{-3}	324.7	175.9	1.846	204811_s_at	calcium channel, voltage-dependent, alpha 2/delta subunit 2; SEQ ID NO:200	CACNA2D2
7.0840×10^{-4}	790.7	419.2	1.886	202725_at	polymerase (RNA) II (DNA directed) polypeptide A, 220kDa ; SEQ ID NO:201	POLR2A
8.8114×10^{-3}	286.6	147	1.95	214741_at	zinc finger protein 131 (clone pHZ-10); SEQ ID NO:202	ZNF131
9.9014×10^{-3}	81	39.9	2.03	212965_at	Hypermethylated in cancer 2; SEQ ID NO:203	HIC2
1.0080×10^{-4}	74.5	35.7	2.087	210641_at	calpain 9 (nCL-4); SEQ ID NO:204	CAPN9
1.5675×10^{-3}	45.2	21.3	2.122	214927_at	integrin, beta-like 1 (with EGF-like repeat domains); SEQ ID NO:205	ITGBL1
4.7900×10^{-5}	309.4	144.3	2.144	40489_at	dentatorubral-pallidolusian atrophy (atrophin-1); SEQ ID NO:206	DRPLA
5.2146×10^{-3}	59.1	27.3	2.165	214650_x_at	myelin oligodendrocyte glycoprotein; SEQ ID NO:207	MOG
7.3571×10^{-3}	635.6	292.9	2.17	214298_x_at	septin 6; SEQ ID NO:208	38236
2.9869×10^{-3}	35.2	16.1	2.186	215392_at	CDNA FLJ14136 fis, clone MAMMA1002744; SEQ ID NO:209	
6.2075×10^{-3}	51.2	23.1	2.216	206644_at	nuclear receptor subfamily 0, group B, member 1; SEQ ID NO:210	NR0B1
1.8113×10^{-3}	140.6	62.3	2.257	214373_at	Clone IMAGE:5171739, mRNA; SEQ ID NO:211	

P-value	Normal	COPD	Fold Change	Probe set ID	Name	Symbol
7.2676×10^{-3}	157.2	66.5	2.364	216248_s_at	nuclear receptor subfamily 4, group A, member 2 ; SEQ ID NO:212	NR4A2
5.9263×10^{-3}	92	38.8	2.371	206229_x_at	paired box gene 2; SEQ ID NO:213	PAX2
6.4400×10^{-5}	69	28.9	2.388	207789_s_at	dipeptidylpeptidase 6; SEQ ID NO:214	DPP6
5.5557×10^{-3}	64.9	26.9	2.413	211343_s_at	collagen, type XIII, alpha 1; SEQ ID NO:215	COL13A1
7.9665×10^{-3}	274.4	112	2.45	216532_x_at	thioredoxin-like 2; SEQ ID NO:216	TXNL2
7.7040×10^{-3}	23.7	9.6	2.469	217440_at	SEQ ID NO:324	
6.1040×10^{-4}	476.5	189.1	2.52	211965_at	zinc finger protein 36, C3H type-like 1; SEQ ID NO:217	ZFP36L1
7.6269×10^{-3}	19.8	7.5	2.64	209700_x_at	phosphodiesterase 4D interacting protein (myomegalin); SEQ ID NO:218	PDE4DIP
5.4251×10^{-3}	100.2	37.6	2.665	205318_at	Kinesin family member 5A ; SEQ ID NO:219	KIF5A
6.5366×10^{-3}	44.6	15.6	2.859	214634_at	Histone 1, H4i; SEQ ID NO:220	HIST1H4I
2.6763×10^{-3}	39.5	13.7	2.883	205069_s_at	Rho GTPase activating protein 26; SEQ ID NO:221	ARHGAP26
6.9970×10^{-3}	37.5	11.9	3.151	213895_at	epithelial membrane protein 1; SEQ ID NO:222	EMP1
1.0189×10^{-3}	50	15.8	3.165	205638_at	brain-specific angiogenesis inhibitor 3; SEQ ID NO:223	BAI3
9.3632×10^{-3}	68.3	20.9	3.268	207287_at	hypothetical protein FLJ14107; SEQ ID NO:224	FLJ14107
4.5021×10^{-3}	42.3	12.5	3.384	204464_s_at	endothelin receptor type A; SEQ ID NO:225	EDNRA
7.6289×10^{-3}	29.2	6.4	4.562	205989_s_at	myelin oligodendrocyte glycoprotein ; SEQ ID NO:207	MOG
3.7723×10^{-3}	44.3	9.6	4.615	210702_s_at	prostaglandin I2 (prostacyclin) synthase; SEQ ID NO:226	PTGIS
4.7300×10^{-3}	69.8	13.9	5.022	206962_x_at	NP220 nuclear protein; SEQ ID NO:227	NP220
2.6600×10^{-5}	22.2	4.3	5.163	207638_at	protease, serine, 7 (enterokinase); SEQ ID NO:228	PRSS7
6.1000×10^{-5}	14.1	1.6	8.812	207107_at	retinal pigment epithelium-specific protein 65kDa ; SEQ ID NO:229	RPE65

Table 2 shows the geometric means of intensities for the genes in both COPD patients (COPD column) and in normal controls (normal column) and provides a fold difference of the mean of intensities (fold change column). Using this information, one can clearly see whether a given gene is upregulated or downregulated in the peripheral blood cells of patients with COPD as compared to the normal control. This table shows the differentially expressed transcripts sorted at $p < 0.01$ sorted by fold-change in the geometric intensities. Therefore, the first

transcript in this table has the highest fold difference between COPD and normal control, and the last transcript in this table has the lowest fold difference, meaning that there is much greater expression in normal controls versus COPD. The genes are identified by name, by probe set identifier and by GenBank Accession numbers. All information associated with the publicly available identifiers and accession numbers in any of the tables described herein, including the nucleic acid sequences of the genes and probes, is incorporated herein by reference in its entirety. The SEQ ID NO's in Table 2 refers to the nucleotide sequence for the coding region of the gene, or, if the entire coding region is not available, whatever fragment of the coding sequence or genomic sequence is available.

Table 3 also reflects a listing of genes identified by the inventors as being differentially expressed in patients with COPD versus normal controls and in this table, results have been grouped into the following main categories: (1) genes that are selectively (i.e., exclusively or uniquely) upregulated in PBMCs of patients with COPD as compared to normal controls (Table 3); and (2) genes that are selectively downregulated in PBMCs of patients with COPD as compared to normal controls (Table 3). Again, the genes are identified by name, by probe set identifier and by GenBank Accession numbers.

Table 3. Genes Selectively Upregulated In Patients With COPD Versus Normal Controls

LocusLink	Gene Name	fold Increase	Parametric p-value	Probe set
3117	major histocompatibility complex, class II, DQ alpha 1; SEQ ID NO:230	90.13	0.0000	213831_at
3119	major histocompatibility complex, class II, DQ beta 1; SEQ ID NO:231	85.08	0.0000	209480_at
3119	major histocompatibility complex, class II, DQ beta 1; SEQ ID NO:231	17.29	0.0001	212999_x_at
3119	major histocompatibility complex, class II, DQ beta 1; SEQ ID NO:231	4.73	0.0001	211654_x_at
995	cell division cycle 25C; SEQ ID NO:232	3.14	0.0029	217010_s_at
5509	protein phosphatase 1, regulatory subunit 3D; SEQ ID NO:3	3.06	0.0022	204555_s_at
645	biliverdin reductase B (flavin reductase (NADPH)); SEQ ID NO:233	2.53	0.0029	202201_at
64170	caspase recruitment domain family, member 9; SEQ ID NO:234	2.49	0.0004	220162_s_at
26272	F-box only protein 4; SEQ ID NO:235	2.39	0.0042	220550_at
54896	hypothetical protein FLJ20320; SEQ ID NO:236	2.36	0.0024	220453_at
7071	TGFB inducible early growth response; SEQ ID NO:237	2.26	0.0036	202393_s_at
9488	phosphatidylinositol glycan, class B; SEQ ID NO:238	1.95	0.0021	214151_s_at
3280	hairy homolog (Drosophila); SEQ ID NO:239	1.93	0.0026	203395_s_at
2736	GLI-Kruppel family member GLI2; SEQ ID NO:240	1.86	0.0012	207034_s_at
7076	tissue inhibitor of metalloproteinase 1 (erythroid potentiating activity, collagenase inhibitor); SEQ ID NO:241	1.83	0.0016	201666_at
54939	hypothetical protein FLJ20452; SEQ ID NO:242	1.82	0.0045	206441_s_at
79166	leukocyte immunoglobulin-like receptor, subfamily A (without TM domain), member 5; SEQ ID NO:243	1.82	0.0024	214561_at

LocusLink	Gene Name	fold Increase	Parametric p-value	Probe set
55954	hypothetical protein LOC55954; SEQ ID NO:244	1.81	0.0050	218752_at
8395	phosphatidylinositol-4-phosphate 5-kinase, type I, beta; SEQ ID NO:245	1.78	0.0029	217477_at
5170	3-phosphoinositide dependent protein kinase-1; SEQ ID NO:246	1.76	0.0035	204524_at
26063	2,4-dienoyl CoA reductase 2, peroxisomal; SEQ ID NO:247	1.76	0.0031	219664_s_at
23129	KIAA0620 protein; SEQ ID NO:248	1.72	0.0035	38671_at
4671	baculoviral IAP repeat-containing 1; SEQ ID NO:249	1.71	0.0043	204861_s_at
2067	excision repair cross-complementing rodent repair deficiency, complementation group 1; SEQ ID NO:250	1.71	0.0033	203720_s_at
10051	SMC4 structural maintenance of chromosomes 4-like 1 (yeast); SEQ ID NO:251	1.68	0.0034	215623_x_at
351	amyloid beta (A4) precursor protein (protease nexin-II, Alzheimer disease); SEQ ID NO:252	1.67	0.0043	200602_at
4166	carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 6; SEQ ID NO:253	1.65	0.0039	221059_s_at
29946	RPA-binding trans-activator; SEQ ID NO:254	1.63	0.0029	219382_at
10981	RAB32, member RAS oncogene family; SEQ ID NO:53	1.63	0.0014	204214_s_at
51097	CGI-49 protein; SEQ ID NO:255	1.61	0.0028	201825_s_at
55359	hypothetical protein DKFZp761P1010; SEQ ID NO:256	1.60	0.0048	221696_s_at
1439	colony stimulating factor 2 receptor, beta, low-affinity (granulocyte-macrophage); SEQ ID NO:102	1.52	0.0047	205159_at
6016	Ras-like without CAAX 1; SEQ ID NO:257	1.51	0.0028	209882_at
3117	major histocompatibility complex, class II, DQ alpha 1; SEQ ID NO:230	90.91	0.0000	203290_at
11273	ataxin 2 related protein; SEQ ID NO:258	5.18	0.0002	207798_s_at
80094	hypothetical protein FLJ14107; SEQ ID NO:224	4.59	0.0001	207287_at
6121	retinal pigment epithelium-specific protein 65kDa; SEQ ID NO:229	4.41	0.0026	207107_at
3119	major histocompatibility complex, class II, DQ beta 1; SEQ ID NO:231	3.26	0.0001	209823_x_at
744	chromosome 11 open reading frame 8; SEQ ID NO:259	3.24	0.0005	215692_s_at
6711	Beta II spectrin [Homo sapiens], mRNA sequence; SEQ ID NO:260	2.98	0.0036	200671_s_at
3119	major histocompatibility complex, class II, DQ beta 1; SEQ ID NO:231	2.75	0.0012	211656_x_at
7539	zinc finger protein 37 homolog (mouse); SEQ ID NO:261	2.67	0.0001	207068_at
57509	AT2 receptor-interacting protein 1; SEQ ID NO:262	2.65	0.0020	212096_s_at
7994	MOZ/CBP [Homo sapiens], mRNA sequence; SEQ ID NO:263	2.65	0.0030	216361_s_at
91695	similar to CGI-96; SEQ ID NO:264	2.55	0.0024	214828_s_at
11118	butyrophilin, subfamily 3, member A2; SEQ ID NO:265	2.40	0.0001	209846_s_at
3655	integrin, alpha 6; SEQ ID NO:266	2.24	0.0013	215177_s_at
55236	hypothetical protein FLJ10808; SEQ ID NO:267	2.16	0.0020	218340_s_at
54941	hypothetical protein FLJ20456; SEQ ID NO:268	2.05	0.0031	207735_at
10753	calpain 9 (nCL-4); SEQ ID NO:204	2.02	0.0025	210641_at
1822	dentatorubral-pallidoluysian atrophy (atrophin-1); SEQ ID NO:206	2.01	0.0036	40489_at
321	amyloid beta (A4) precursor protein-binding, family A, member 2 (X11-like); SEQ ID NO:269	1.96	0.0004	209870_s_at
493	ATPase, Ca ⁺⁺ transporting, plasma membrane 4; SEQ ID NO:270	1.91	0.0049	205410_s_at
3631	inositol polyphosphate-4-phosphatase, type I, 107kDa; SEQ ID NO:271	1.89	0.0012	204553_x_at
7556	zinc finger protein 10 (KOX 1); SEQ ID NO:272	1.87	0.0039	216350_s_at
9991	ROD1 regulator of differentiation 1 (S. pombe); SEQ ID NO:273	1.86	0.0019	214698_at
65982	hypothetical protein FLJ12895; SEQ ID NO:274	1.85	0.0050	218312_s_at

LocusLink	Gene Name	fold Increase	Parametric p-value	Probe set
8718	tumor necrosis factor receptor superfamily, member 25; SEQ ID NO:275	1.81	0.0035	211841 s at
55769	zinc finger protein 83 (HPF1); SEQ ID NO:276	1.79	0.0009	221645 s at
51303	FK506 binding protein 11, 19 kDa; SEQ ID NO:277	1.77	0.0024	219118 at
3423	iduronate 2-sulfatase (Hunter syndrome); SEQ ID NO:278	1.75	0.0012	202438 x at
3655	integrin, alpha 6; SEQ ID NO:266	1.75	0.0037	201656 at
1565	cytochrome P450, subfamily IID (debrisoquine, sparteine, etc., -metabolizing), polypeptide 6; SEQ ID NO:279	1.67	0.0007	207498 s at
80067	hypothetical protein FLJ13096; SEQ ID NO:280	1.65	0.0029	220172 at
51281	testis specific ankyrin-like protein 1; SEQ ID NO:281	1.65	0.0042	210486 at
613	breakpoint cluster region; SEQ ID NO:282	1.62	0.0022	217223 s at
23176	likely ortholog of mouse septin 8; SEQ ID NO:283	1.62	0.0024	208999 at
1503	CTP synthase; SEQ ID NO:284	1.61	0.0022	202613 at
7571	zinc finger protein 23 (KOX 16); SEQ ID NO:285	1.58	0.0047	213934 s at
9666	KIAA0675 gene product; SEQ ID NO:286	1.57	0.0038	213186 at
9112	metastasis associated 1; SEQ ID NO:287	1.52	0.0041	211783 s at
25865	protein kinase D2; SEQ ID NO:288	1.50	0.0033	38269 at
5567	protein kinase, cAMP-dependent, catalytic, beta; SEQ ID NO:289	1.45	0.0040	202741 at

Tables 4 and 5 also reflect a listing of genes identified by the inventors as being differentially expressed in patients with COPD versus normal controls. These tables show the differentially expressed transcripts sorted at $p < 0.005$ sorted by fold-change in the geometric intensities. PSEM refers to "past smoker emphysema" and is equivalent to the category of "COPD" as set forth in Table 2. NSNL refers to "non-smoker normal" and is equivalent to the category of "Normal" as set forth in Table 2. Again, the genes are identified by name, by probe set identifier and by GenBank Accession numbers. In Tables 4 and 5, 22277 genes were filtered, and 13909 of these passed filtering criteria. A two-sample T-test (with randomized variance model) was used. The Multivariate Permutations test was computed based on 1000 permutations. The nominal significance level of each univariate test was 0.005. The confidence level of false discovery rate assessment was 50%. The maximum allowed number of false-positive genes was 10, and the maximum allowed proportion of false-positive genes: 0.1. In Table 4, the number of genes significant at 0.005 level of the univariate test was 66, and probability of getting at least 66 genes significant by chance (at the 0.005 level) if there are no real differences between the classes was 0.283. In Table 5, the number of genes significant at 0.005 level of the univariate test was 77, and the probability of getting at least 77 genes significant by chance (at the 0.005 level) if there are no real differences between the classes: 0.285. In Table 4, the predicted number of false discoveries among the first 10 genes is 10, and the predicted proportion of false discoveries among the first 0 genes is 10%. In Table 5, the

predicted number of false discoveries among the first 21 genes is 10, and the predicted proportion of false discoveries among the first 6 genes is 10%.

Table 4. Genes identified as being differentially expressed in patients with COPD versus normal controls.

Parametric p-value	Mean of intensities in PSEM	Mean of intensities in NSNL	Fold difference of geom means	Probe set	Description	Anno-tations	UG cluster	Gene symbol
0.0001576	24.168	1231.986	0.02	203290_at	major histocompatibility complex, class II, DQ alpha 1; SEQ ID NO:230	Info	Hs.198253	HLA-DQA1
0.0001934	12.268	57.917	0.212	207798_s at	ataxin 2 related protein; SEQ ID NO:258	Info	Hs.43509	A2LP
0.0002381	11.835	37.215	0.318	215692_s at	chromosome 11 open reading frame 8; SEQ ID NO:259	Info	Hs.46638	C11orf8
0.0003165	1288.476	2856.959	0.451	209846_s at	butyrophilin, subfamily 3, member A2; SEQ ID NO:265	Info	Hs.87497	BTN3A2
0.000467	96.222	50.618	1.901	207034_s at	GLI-Kruppel family member GLI2; SEQ ID NO:240	Info	Hs.111867	GLI2
0.0004852	71.787	117.45	0.611	207498_s at	cytochrome P450, subfamily IID (debrisoquine, sparteine, etc., -metabolizing), polypeptide 6; SEQ ID NO:279	Info	Hs.333497	CYP2D6
0.0008842	146.181	316.905	0.461	218340_s at	hypothetical protein FLJ10808; SEQ ID NO:267	Info	Hs.59838	FLJ10808
0.0013	140.533	240.396	0.585	221645_s at	zinc finger protein 83 (HPF1); SEQ ID NO:276	Info	Hs.305953	ZNF83
0.0013175	2.623	13.073	0.201	207107_at	retinal pigment epithelium-specific protein 65kDa; SEQ ID NO:229	Info	Hs.2133	RPE65
0.0014033	37.553	74.637	0.503	210641_at	calpain 9 (nCL-4); SEQ ID NO:204	Info	Hs.113292	CAPN9
0.0014101	18.841	71.508	0.263	207287_at	hypothetical protein FLJ14107; SEQ ID NO:224	Info	Hs.287624	FLJ14107
0.0014649	1054.882	1580.62	0.667	38269_at	protein kinase D2; SEQ ID NO:288	Info	Hs.91146	PKD2
0.0016367	34.683	12.334	2.812	216814_at	; SEQ ID NO:290			
0.0016372	102.802	163.113	0.63	208999_at	likely ortholog of mouse septin 8; SEQ ID NO:283	Info	Hs.80712	37872

Parametric p-value	Mean of intensities in PSEM	Mean of intensities in NSNL	Fold difference of geom means	Probe set	Description	Annotations	UG cluster	Gene symbol
0.0016477	1737.224	1104.834	1.572	204214_s at	RAB32, member RAS oncogene family; SEQ ID NO:53	Info	Hs.32217	RAB32
0.0016522	82.456	134.398	0.614	220172_at	hypothetical protein FLJ13096; SEQ ID NO:280	Info	Hs.287549	FLJ13096
0.0017036	193.937	349.942	0.554	204553_x at	inositol polyphosphate-4-phosphatase, type I, 107kDa; SEQ ID NO:271	Info	Hs.32944	INPP4A
0.0017859	239.102	105.958	2.257	220162_s at	caspase recruitment domain family, member 9; SEQ ID NO:234	Info	Hs.271815	CARD9
0.0021283	156.556	54.631	2.866	204555_s at	protein phosphatase 1, regulatory subunit 3D; SEQ ID NO:3	Info	Hs.42215	PPP1R3D
0.002212	74.261	121.622	0.611	210486_at	testis specific ankyrin-like protein 1; SEQ ID NO:281	Info	Hs.122275	FLJ20499
0.0023461	175.047	90.699	1.93	215063_x at	Homo sapiens mRNA; cDNA DKFZp547B026 (from clone DKFZp547B026), mRNA sequence; SEQ ID NO:291	Info	Hs.283862	
0.002444	34.835	88.105	0.395	215135_at	aspartyl aminopeptidase; SEQ ID NO:	Info	Hs.258551	DNPEP
0.0024685	28.529	60.877	0.469	208297_s at	ecotropic viral integration site 5; SEQ ID NO:292	Info	Hs.179747	EVI5
0.0024778	160.517	250.224	0.641	202613_at	CTP synthase; SEQ ID NO:284	Info	Hs.251871	CTPS
0.0024821	20.678	52.546	0.394	213703_at	Unnamed protein product [Homo sapiens], mRNA sequence; SEQ ID NO:293	Info	Hs.349607	
0.0024932	310.282	529.848	0.586	219118_at	FK506 binding protein 11, 19 kDa; SEQ ID NO:277	Info	Hs.24048	FKBP11
0.0026417	11938.343	7624.437	1.566	211750_x at	tubulin alpha 6; SEQ ID NO:92	Info	Hs.406578	TUBA6
0.0027147	40.851	17.527	2.331	220550_at	F-box only protein 4; SEQ ID NO:235	Info	Hs.49210	FBXO4
0.0027235	55.264	136.33	0.405	212096_s at	AT2 receptor-interacting protein 1; SEQ ID NO:262	Info	Hs.7946	ATIP1

Parametric p-value	Mean of intensities in PSEM	Mean of intensities in NSNL	Fold difference of geom means	Probe set	Description	Anno-tations	UG cluster	Gene symbol
0.0028465	73.759	49.848	1.48	206533_at	cholinergic receptor, nicotinic, alpha polypeptide 5; SEQ ID NO:294	Info	Hs.1614	CHRNA5
0.0028577	184.921	105.862	1.747	205247_at	Notch homolog 4 (Drosophila); SEQ ID NO:295	Info	Hs.11689	NOTCH4
0.003006	195.897	284.134	0.689	212218_s at	fatty acid synthase; SEQ ID NO:296	Info	Hs.83190	FASN
0.0030551	1186.231	803.474	1.476	218317_x at	hypothetical protein MGC5178; SEQ ID NO:297	Info	Hs.326067	MGC5178
0.0030663	12.057	34.887	0.346	213895_at	epithelial membrane protein 1; SEQ ID NO:222	Info	Hs.79368	EMP1
0.0030891	25.413	44.76	0.568	214569_at	interferon, alpha 5 check 197; SEQ ID NO:191	Info	Hs.37113	IFNA5
0.0030957	293.646	152.269	1.928	219664_s at	2,4-dienoyl CoA reductase 2, peroxisomal; SEQ ID NO:247	Info	Hs.15898	DECR2
0.0031333	116.894	194.042	0.602	205395_s at	MRE11 meiotic recombination 11 homolog A (S. cerevisiae); SEQ ID NO:298	Info	Hs.20555	MRE11A
0.0032464	1344.768	2316.231	0.581	209829_at	chromosome 6 open reading frame 32; SEQ ID NO:299	Info	Hs.101359	C6orf32
0.003251	730.175	326.175	2.239	AFFX-BioB-M at				
0.0033265	65.856	116.546	0.565	205362_s at	prefoldin 4; SEQ ID NO:300	Info	Hs.91161	PFDN4
0.0033531	127.961	238.935	0.536	205410_s at	ATPase, Ca ⁺⁺ transporting, plasma membrane 4; SEQ ID NO:270	Info	Hs.343522	ATP2B4
0.0033631	43.633	75.778	0.576	211220_s at	heat shock transcription factor 2; SEQ ID NO:301	Info	Hs.158195	HSF2
0.0034281	1856.035	685.466	2.708	210075_at	hypothetical protein LOC51257; SEQ ID NO:302	Info	Hs.132744	LOC51257
0.003443	299.05	175.87	1.7	204524_at	3-phosphoinositide dependent protein kinase-1; SEQ ID NO:246	Info	Hs.154729	PDPK1
0.0034777	155.544	283.521	0.549	206914_at	class-I MHC-restricted T cell associated molecule; SEQ ID NO:303	Info	Hs.159523	CRTAM

Parametric p-value	Mean of intensities in PSEM	Mean of intensities in NSNL	Fold difference of geom means	Probe set	Description	Annotations	UG cluster	Gene symbol
0.003539	2993.317	1393.766	2.148	202393_s at	TGFB inducible early growth response; SEQ ID NO:237	Info	Hs.82173	TIEG
0.0035561	3233.094	2169.713	1.49	205159_at	colony stimulating factor 2 receptor, beta, low-affinity (granulocyte-macrophage); SEQ ID NO:102	Info	Hs.285401	CSF2RB
0.0036176	10485.038	6555.809	1.599	221059_s at	carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 6; SEQ ID NO:253	Info	Hs.157439	CHST6
0.0036459	495.018	269.218	1.839	214151_s at	phosphatidylinositol glycan, class B; SEQ ID NO:238	Info	Hs.247118	PIGB
0.0037128	742.272	509.402	1.457	209882_at	Ras-like without CAAX 1; SEQ ID NO:257	Info	Hs.96038	RIT1
0.0037852	1118.073	523.085	2.137	204418_x at	glutathione S-transferase M2 (muscle); SEQ ID NO:304	Info	Hs.279837	GSTM2
0.0038778	102.203	160.646	0.636	210796_x at	sialic acid binding Ig-like lectin 6; SEQ ID NO:305	Info	Hs.117992	SIGLEC6
0.0039155	1371.432	854.809	1.604	204119_s at	adenosine kinase; SEQ ID NO:306	Info	ADK	
0.004225	174.326	265.575	0.656	206574_s at	protein tyrosine phosphatase type IVA, member 3; SEQ ID NO:307	Info	Hs.43666	PTP4A3
0.0043417	86.042	47.378	1.816	203395_s at	hairy homolog (Drosophila); SEQ ID NO:239	Info	Hs.250666	HRY
0.0043563	86.758	163.229	0.532	206974_at	chemokine (C-X-C motif) receptor 6; SEQ ID NO:308	Info	Hs.34526	CXCR6
0.0044602	339.664	220.304	1.542	219122_s at	hypothetical protein FLJ20546; SEQ ID NO:309	Info	Hs.353090	FLJ20546
0.0044638	73.533	45.625	1.612	215623_x at	SMC4 structural maintenance of chromosomes 4-like 1 (yeast); SEQ ID NO:251	Info	Hs.50758	SMC4L1
0.0045287	5236.698	3053.492	1.715	201666_at	tissue inhibitor of metalloproteinase 1 (erythroid potentiating activity, collagenase inhibitor); SEQ ID NO:241	Info	Hs.433425	TIMP1

Parametric p-value	Mean of intensities in PSEM	Mean of intensities in NSNL	Fold difference of geom means	Probe set	Description	Annotations	UG cluster	Gene symbol
0.0045448	23.022	52.709	0.437	207068_at	zinc finger protein 37 homolog (mouse); SEQ ID NO:261	Info	Hs.150406	ZFP37
0.0045651	155.943	62.982	2.476	216833_x_at	glycophorin B (includes Ss blood group); SEQ ID NO:310	Info	Hs.372513	GYPB
0.0047352	2467.916	61.355	40.224	213831_at	major histocompatibility complex, class II, DQ alpha 1; SEQ ID NO:230	Info	Hs.198253	HLA-DQA1
0.0047656	81.254	55.79	1.456	215629_s_at	BCMS upstream neighbor-like; SEQ ID NO:311	Info	Hs.351281	BCMSUNL
0.004873	121.362	63.316	1.917	220641_at	NADPH oxidase, EF hand calcium-binding domain 5; SEQ ID NO:312	Info	Hs.160199	NOX5
0.0048741	42.334	66.738	0.634	210286_s_at	solute carrier family 4, sodium bicarbonate cotransporter, member 7; SEQ ID NO:313	Info	Hs.132904	SLC4A7
0.004877	854.828	353.837	2.416	204550_x_at	glutathione S-transferase M1; SEQ ID NO:314	Info	Hs.301961	GSTM1

Table 5. Genes identified as being differentially expressed in patients with COPD versus normal controls.

Parametric p-value	Mean of intensities in PSEM	Mean of intensities in NSNL	Fold difference of geom means	Probe set	Description	UG cluster	Gene symbol
p < 1e-07	24.168	2126.366	0.011	203290_at	major histocompatibility complex, class II, DQ alpha 1; SEQ ID NO:230	Hs.198253	HLA-DQA1
9.00E-07	2467.916	27.383	90.126	213831_at	major histocompatibility complex, class II, DQ alpha 1; SEQ ID NO:230	Hs.198253	HLA-DQA1
1.87E-05	1509.817	17.745	85.084	209480_at	major histocompatibility complex, class II, DQ beta 1; SEQ ID NO:231	Hs.73931	HLA-DQB1
5.50E-05	1288.476	3096.248	0.416	209846_s_at	butyrophilin, subfamily 3, member A2; SEQ ID NO:265	Hs.87497	BTN3A2
6.68E-05	18.841	86.572	0.218	207287_at	hypothetical protein FLJ14107; SEQ ID NO:224	Hs.287624	FLJ14107
8.28E-05	1029.073	3349.976	0.307	209823_x_at	major histocompatibility complex, class II, DQ beta 1; SEQ ID NO:231	Hs.73931	HLA-DQB1
0.0001021	23.022	61.354	0.375	207068_at	zinc finger protein 37 homolog (mouse); SEQ ID NO:261	Hs.150406	ZFP37

Parametric p-value	Mean of intensities in PSEM	Mean of intensities in NSNL	Fold difference of geom means	Probe set	Description	UG cluster	Gene symbol
0.0001105	5277.87	1116.924	4.725	211654_x at	major histocompatibility complex, class II, DQ beta 1; SEQ ID NO:231	Hs.73931	HLA-DQB1
0.0001106	598.891	34.639	17.29	212999_x at	major histocompatibility complex, class II, DQ beta 1; SEQ ID NO:231	Hs.73931	HLA-DQB1
0.0001255	34.683	10.679	3.248	216814_at	; SEQ ID NO:290		
0.0001685	12.268	63.498	0.193	207798_s at	ataxin 2 related protein; SEQ ID NO:258	Hs.43509	A2LP
0.0003715	239.102	96.001	2.491	220162_s at	caspace recruitment domain family, member 9; SEQ ID NO:234	Hs.271815	CARD9
0.0003822	326.747	641.328	0.509	209870_s at	amyloid beta (A4) precursor protein-binding, family A, member 2 (X11-like); SEQ ID NO:	Hs.26468	APBA2
0.0004507	11.835	38.314	0.309	215692_s at	chromosome 11 open reading frame 8; SEQ ID NO:259	Hs.46638	C11orf8
0.0007448	71.787	119.867	0.599	207498_s at	cytochrome P450, subfamily IID (debrisoquine, sparteine, etc., -metabolizing), polypeptide 6; SEQ ID NO:279	Hs.333497	CYP2D6
0.000888	140.533	251.152	0.56	221645_s at	zinc finger protein 83 (HPF1); SEQ ID NO:276	Hs.305953	ZNF83
0.0011876	196.535	343.84	0.572	202438_x at	Homo sapiens mRNA; cDNA DKFZp434G012 (from clone DKFZp434G012), mRNA sequence; SEQ ID NO:315	Hs.303154	
0.001225	1240.817	3406.64	0.364	211656_x at	major histocompatibility complex, class II, DQ beta 1; SEQ ID NO:231	Hs.73931	HLA-DQB1
0.0012391	193.937	366.926	0.529	204553_x at	inositol polyphosphate-4-phosphatase, type I, 107kDa; SEQ ID NO:271	Hs.32944	INPP4A
0.0012462	96.222	51.635	1.864	207034_s at	GLI-Kruppel family member GLI2; SEQ ID NO:240	Hs.111867	GLI2
0.0012933	399.647	894.951	0.447	215177_s at	integrin, alpha 6; SEQ ID NO:266	Hs.227730	ITGA6
0.0014275	1737.224	1068.632	1.626	204214_s at	RAB32, member RAS oncogene family; SEQ ID NO:53	Hs.32217	RAB32
0.001619	5236.698	2859.16	1.832	201666_at	tissue inhibitor of metalloproteinase 1 (erythroid potentiating activity, collagenase inhibitor); SEQ ID NO:241	Hs.433425	TIMP1
0.0018609	259.179	480.63	0.539	214698_at	ROD1 regulator of differentiation 1 (S. pombe); SEQ ID NO:273	Hs.374634	ROD1
0.0020119	146.181	315.775	0.463	218340_s at	hypothetical protein FLJ10808; SEQ ID NO:267	Hs.59838	FLJ10808
0.002022	55.264	146.737	0.377	212096_s at	AT2 receptor-interacting protein 1; SEQ ID NO:262	Hs.7946	ATIP1
0.0021169	495.018	253.682	1.951	214151_s at	phosphatidylinositol glycan, class B; SEQ ID NO:238	Hs.247118	PIGB
0.0021708	160.517	258.718	0.62	202613_at	CTP synthase; SEQ ID NO:284	Hs.251871	CTPS
0.0021732	156.556	51.186	3.059	204555_s at	protein phosphatase 1, regulatory subunit 3D; SEQ ID NO:3	Hs.42215	PPP1R3D
0.0022344	236.002	382.391	0.617	217223_s at	breakpoint cluster region; SEQ ID NO:282	Hs.234799	BCR
0.0023812	310.282	549.738	0.564	219118_at	FK506 binding protein 11, 19 kDa; SEQ ID NO:277	Hs.24048	FKBP11

Parametric p-value	Mean of intensities in PSEM	Mean of intensities in NSNL	Fold difference of geom means	Probe set	Description	UG cluster	Gene symbol
0.0023874	49.48	27.208	1.819	214561_at	leukocyte immunoglobulin-like receptor, subfamily A (without TM domain), member 5; SEQ ID NO:316	Hs.202680	ILT10
0.0023948	134.325	56.839	2.363	220453_at	hypothetical protein FLJ20320; SEQ ID NO:236	Hs.263081	FLJ20320
0.0024082	72.289	184.59	0.392	214828_s at	similar to CGI-96; SEQ ID NO:264	Hs.387850	dJ222E13.2
0.0024152	102.802	166.207	0.619	208999_at	likely ortholog of mouse septin 8; SEQ ID NO:283	Hs.80712	37872
0.002537	37.553	75.664	0.496	210641_at	calpain 9 (nCL-4); SEQ ID NO:204	Hs.113292	CAPN9
0.0026238	2.623	11.56	0.227	207107_at	retinal pigment epithelium-specific protein 65kDa; SEQ ID NO:229	Hs.2133	RPE65
0.0026369	86.042	44.701	1.925	203395_s at	hairly homolog (Drosophila); SEQ ID NO:239	Hs.250666	HRY
0.0027798	359.43	223.624	1.607	201825_s at	CGI-49 protein; SEQ ID NO:255	Hs.238126	CGI-49
0.0027827	742.272	492.82	1.506	209882_at	Ras-like without CAAX 1; SEQ ID NO:257	Hs.96038	RIT1
0.0028527	499.986	306.419	1.632	219382_at	RPA-binding trans-activator; SEQ ID NO:254	Hs.169138	RBT1
0.0028612	1921.032	760.74	2.525	202201_at	biliverdin reductase B (flavin reductase (NADPH)); SEQ ID NO:233	Hs.76289	BLVRB
0.0028921	59.806	19.072	3.136	217010_s at	cell division cycle 25C; SEQ ID NO:232	Hs.656	CDC25C
0.002944	134.984	75.859	1.779	217477_at	phosphatidylinositol-4-phosphate 5-kinase, type I, beta; SEQ ID NO:245	Hs.78406	PIP5K1B
0.0029449	82.456	135.921	0.607	220172_at	hypothetical protein FLJ13096; SEQ ID NO:280	Hs.287549	FLJ13096
0.0029522	70.686	187.302	0.377	216361_s at	MOZ/CBP [Homo sapiens], mRNA sequence; SEQ ID NO:263	Hs.381462	
0.0029685	730.175	308.016	2.371	AFFX-BioB-M_at			
0.0030822	293.646	166.974	1.759	219664_s at	2,4-dienoyl CoA reductase 2, peroxisomal; SEQ ID NO:247	Hs.15898	DECR2
0.0031083	130.467	267.467	0.488	207735_at	hypothetical protein FLJ20456; SEQ ID NO:268	Hs.272800	FLJ20456
0.0032625	643.678	376.097	1.711	203720_s at	excision repair cross-complementing rodent repair deficiency, complementation group 1 (includes overlapping antisense sequence); SEQ ID NO:317	Hs.59544	ERCC1
0.0032992	1054.882	1579.764	0.668	38269_at	protein kinase D2; SEQ ID NO:288	Hs.91146	PKD2
0.0034086	73.533	43.783	1.679	215623_x at	SMC4 structural maintenance of chromosomes 4-like 1 (yeast); SEQ ID NO:251	Hs.50758	SMC4L1
0.0035025	129.018	233.598	0.552	211841_s at	tumor necrosis factor receptor superfamily, member 25; SEQ ID NO:275	Hs.180338	TNFRSF25
0.0035194	299.05	169.85	1.761	204524_at	3-phosphoinositide dependent protein kinase-1; SEQ ID NO:246	Hs.154729	PDPK1
0.0035245	1015.829	591.253	1.718	38671_at	KIAA0620 protein; SEQ ID NO:248	Hs.301685	KIAA0620
0.0035525	49.828	148.294	0.336	200671_s at	Beta II spectrin [Homo sapiens], mRNA sequence; SEQ ID NO:318	Hs.348723	

Parametric p-value	Mean of intensities in PSEM	Mean of intensities in NSNL	Fold difference of geom means	Probe set	Description	UG cluster	Gene symbol
0.0036142	170.654	342.906	0.498	40489 at	dentatorubral-pallidoluysian atrophy (atrophin-1); SEQ ID NO:206	Hs.169488	DRPLA
0.0036417	2993.317	1327.645	2.255	202393 s at	TGFB inducible early growth response; SEQ ID NO:237	Hs.82173	TIEG
0.0036778	313.925	548.906	0.572	201656 at	integrin, alpha 6; SEQ ID NO:266	Hs.227730	ITGA6
0.0038219	153.659	241.602	0.636	213186 at	KIAA0675 gene product; SEQ ID NO:286	Hs.165662	KIAA0675
0.0038982	45.302	84.893	0.534	216350 s at	zinc finger protein 10 (KOX 1); SEQ ID NO:272	Hs.104115	ZNF10
0.003906	10485.038	6365.617	1.647	221059 s at	carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 6; SEQ ID NO:253	Hs.157439	CHST6
0.0039763	3372.867	4896.314	0.689	202741 at	protein kinase, cAMP-dependent, catalytic, beta; SEQ ID NO:289	Hs.87773	PRKACB
0.004076	380.164	576.944	0.659	211783 s at	metastasis associated 1; SEQ ID NO:287	Hs.101448	MTA1
0.0041776	40.851	17.09	2.39	220550 at	F-box only protein 4; SEQ ID NO:235	Hs.49210	FBXO4
0.0042365	74.261	122.265	0.607	210486 at	testis specific ankyrin-like protein 1; SEQ ID NO:281	Hs.122275	FLJ20499
0.0042898	279.751	163.321	1.713	204861 s at	baculoviral IAP repeat-containing 1; SEQ ID NO:249	Hs.79019	BIRC1
0.0043342	848.119	506.847	1.673	200602 at	amyloid beta (A4) precursor protein (protease nexin-II, Alzheimer disease); SEQ ID NO:	Hs.177486	APP
0.0044481	599.744	361.75	1.658	216336 x at	SEQ ID NO:319		
0.0045454	381.74	209.794	1.82	206441 s at	hypothetical protein FLJ20452; SEQ ID NO:242	Hs.351327	FLJ20452
0.0046369	175.047	90.607	1.932	215063 x at	Homo sapiens mRNA; cDNA DKFZp547B026 (from clone DKFZp547B026), mRNA sequence; SEQ ID NO:	Hs.283862	
0.0046926	3233.094	2128.235	1.519	205159 at	colony stimulating factor 2 receptor, beta, low-affinity (granulocyte-macrophage); SEQ ID NO:102	Hs.285401	CSF2RB
0.004715	101.294	160.511	0.631	213934 s at	zinc finger protein 23 (KOX 16); SEQ ID NO:285	Hs.376810	ZNF23
0.0048349	62.857	39.413	1.595	221696 s at	hypothetical protein DKFZp761P1010; SEQ ID NO:256	Hs.24979	DKFZp761P1010
0.0049348	127.961	244.126	0.524	205410 s at	ATPase, Ca ⁺⁺ transporting, plasma membrane 4; SEQ ID NO:270	Hs.343522	ATP2B4
0.0049769	326.181	180.093	1.811	218752 at	hypothetical protein LOC55954; SEQ ID NO:244	Hs.38628	LOC55954
0.0049818	361.484	669.276	0.54	218312 s at	hypothetical protein FLJ12895; SEQ ID NO:274	Hs.235390	FLJ12895

It should be noted that certain genes appear more than once in the tables provided herein, and in some cases, a gene may appear by name in both the "upregulated" and the "downregulated" category. This is because well-annotated genes often have multiple probe sets that one can use to identify the gene, and also because various isoforms of certain genes may be

included, where there is some variation in the isotype sequence that is reflected by the various probe sets on the microarray chip (*i.e.*, the probe sets are capable of differentiating between different isotypes of the same gene). As such, one isotype may be upregulated as compared to normal controls, where a second isotype may be down regulated as compared to normal controls.

Accordingly, in one embodiment of the present invention, the genes identified as being regulated (upregulated or downregulated) in PBMCs of patients with COPD can be used as endpoints or markers (also called "biomarkers") in a diagnostic or prognostic assay for COPD. The biomarkers include any of the genes listed in any of the tables presented herein (*e.g.*, Tables 2-5). Diagnostic assays include assays that determine whether a patient has overt COPD or preclinical stage COPD. Prognostic assays can be used to stage a patient's development of COPD, predict a patient's outcome or disease progression, and/or monitor the effectiveness of various treatment protocols on COPD.

The term "biomarker" as used herein can refer to an endpoint gene described herein or to the protein encoded by that gene. In addition, the term "biomarker" can be generally used to refer to any portion of such a gene or protein that can identify or correlate with the full-length gene or protein, for example, in an assay of the invention. According to the present invention, an "endpoint gene" or "biomarker gene" is any gene, the expression of which is regulated (up or down) in a patient with a condition as compared to a normal control. Selected sets of one, two, three, and more preferably several more of the genes of this invention (up to the number equivalent to all of the genes, including any intervening number, in whole number increments, *e.g.*, 1, 2, 3, 4, 5, 6...) can be used as end-points for rapid diagnostics or prognostics for COPD. Preferably, larger numbers of the genes identified in any one or more of Tables 2-5 are used in an assay of the invention (*e.g.*, at least 10 genes or more), since the accuracy of the assay improves as the number of genes screened increases.

According to the present invention, the method includes the step of detecting the expression of at least one, and preferably more than one (*e.g.*, 2, 3, 4, 5, 6,...and so on, in increments of whole numbers up to all of the genes) of the genes that have now been shown to be selectively regulated in PBMCs of patients with COPD by the present inventors. As used herein, the term "expression", when used in connection with detecting the expression of a gene of the present invention, can refer to detecting transcription of the gene and/or to detecting translation of the gene. To detect expression of a gene refers to the act of actively determining whether a gene is expressed or not. This can include determining whether the gene expression is upregulated as compared to a control, downregulated as compared to a control, or

substantially unchanged as compared to a control. Therefore, the step of detecting expression does not require that expression of the gene actually is upregulated or downregulated, but rather, can also include detecting no expression of the gene or detecting that the expression of the gene has not changed or is not different (i.e., detecting no significant expression of the gene or no significant change in expression of the gene as compared to a control).

The present method includes the step of detecting the expression of at least one gene that is selectively regulated in PBMCs of a patient with COPD. In a preferred embodiment, the step of detecting includes detecting the expression of at least 2 genes, and preferably at least 3 genes, and more preferably at least 4 genes, and more preferably at least 5 genes, and more preferably at least 6 genes, and more preferably at least 7 genes, and more preferably at least 8 genes, and more preferably at least 9 genes, and more preferably at least 10 genes, and more preferably at least 11 genes, and more preferably at least 12 genes, and more preferably at least 13 genes, and more preferably at least 14 genes, and more preferably at least 15 genes, and more preferably at least 20 genes, and more preferably at least 25 genes, and more preferably at least 50 genes, and more preferably at least 75 genes, and more preferably at least 100 genes, and so on, in whole integer increments (i.e., 1, 2, 3,...10, 11, 12,...35, 36, 37,...56, 57, 58,...98, 99, 100,...106), up to detecting expression of all of the genes that can be used to detect COPD as disclosed herein. Analysis of a number of genes greater than one can be accomplished simultaneously, sequentially, or cumulatively. As discussed above, it is preferred that several (e.g., at least 10) and up to most or all of the genes be detected in the present methods, as the accuracy of the method improves as the number of genes detected increases. However, it is to be understood that in some circumstances, it may be desirable and sufficient to detect the expression of only one or a few genes.

In the diagnostic or prognostic method of the present invention, the gene(s) to be detected are preferably selected from the genes described in any one or more of Tables 2-5 (i.e., Table 2, Table 3, Table 4 or Table 5, or any combination thereof). These tables have been discussed above in detail and disclose genes that the present inventors have discovered to be selectively regulated in the PBMCs of patients with COPD. More specifically, these tables disclose the manner in which the genes are regulated (e.g., upregulated or downregulated) in a patient with COPD as compared to a normal control.

It is to be understood that the organization of various genes into the present tables is for purposes of illustrating various experimental data described in the Examples section. The selection of genes to be detected in any given method can include any one or more of the genes in any one or more of Tables 2-5, and can include the detection of any combination of two or

more of the genes in any one or more of Tables 2-5, and preferably includes the detection of any combination of multiple genes (*e.g.*, at least 3, 4, 5, 6, ...up to all of the genes) in any one or more of Tables 2-5. It is not mandatory that a given assay be restricted to the detection of all of the various genes in a single table, or to at least one gene in each table. In addition, one may
5 choose also to detect other genes that are believed to be useful in the evaluation of a patient for COPD, and therefore, the present method is not limited exclusively to detection of the genes identified herein, although the invention is primarily directed to the detection of one or more of these genes and includes the detection of at least one or more of these genes. In addition, provided with this disclosure, one of skill in the art may proceed to identify additional genes
10 that are differentially regulated in the PBMCs of patients with COPD, and detection of any of such genes may be used in the methods of the present invention, including in combination with detection of any of the genes disclosed herein. Indeed, the present inventors have now provided a powerful method to detect and evaluate biomarkers for COPD and have also provided data demonstrating the application of such technology.

15 Given the knowledge of the genes regulated in COPD according to the present invention, one of skill in the art will be able to select one or more genes (at least one gene, and preferably, two, three, four, or any number of additional genes) to detect in a method of the present invention, and the selection of the one or more genes can be determined based on the preferences of the person using the assays described herein.

20 In one aspect, it may be desirable to preferentially select those genes for detection that are particularly highly regulated in patients with COPD in that they display the largest increases or decreases in expression levels in patients as compared to normal controls or as compared to the other form of COPD. The detection of such genes can be advantageous because the endpoint may be more clear and require less quantitation. The relative expression levels of the
25 genes identified in the present invention are listed in the tables.

According to the present invention, a "baseline" or "control" can include a normal or negative control and/or a disease or positive control, against which a test level of gene expression can be compared. Therefore, it can be determined, based on the control or baseline level of gene expression, whether a sample to be evaluated for COPD has a measurable
30 difference or substantially no difference in gene expression, as compared to the baseline level. In one aspect, the baseline control is indicative of the level of gene expression as expected in the PBMCs of a normal individual (*e.g.*, healthy individual, negative control, or non-COPD patient). Therefore, the term "negative control" used in reference to a baseline level of gene expression typically refers to a baseline level of expression from a population of individuals

which is believed to be normal (*i.e.*, not having or developing COPD). In some embodiments of the invention, it may also be useful to compare the gene expression in a test sample of PBMCs to a baseline that has previously been established from a patient or population of patients with COPD. Such a baseline level, also referred to herein as a "positive control", refers to a level of gene expression established in PBMCs from one or preferably a population of individuals who had been positively diagnosed with COPD.

In one embodiment, when the goal is to monitor the progression or regression of COPD in a patient, for example, to monitor the efficacy of treatment of the disease or to determine whether a patient that appears to be predisposed to the disease begins to develop the disease, one baseline control can include the measurements of gene expression in a sample of PBMCs from the patient that was taken from a prior test in the same patient. In this embodiment, a new sample is evaluated periodically (*e.g.*, at annual or more regular physicals), and any changes in gene expression in the patient PBMCs as compared to the prior measurement and most typically, also with reference to the above-described normal and/or positive controls, are monitored. Monitoring of a patient's PBMC gene expression profile can be used by the clinician to prescribe or modify treatment for the patient based on whether any differences in gene expression in the PBMCs is indicated.

In a preferred embodiment, the control or baseline levels of gene expression are obtained from PBMCs collected from "matched individuals". According to the present invention, the phrase "matched individuals" refers to a matching of the control individuals on the basis of one or more characteristics, such as gender, age, race, or any relevant biological or sociological factor that may affect the baseline of the control individuals and the patient (*e.g.*, preexisting conditions, consumption of particular substances, levels of other biological or physiological factors). The number of matched individuals from whom control samples must be obtained to establish a suitable control level (*e.g.*, a population) can be determined by those of skill in the art, but should be statistically appropriate to establish a suitable baseline for comparison with the patient to be evaluated (*i.e.*, the test patient). The values obtained from the control samples are statistically processed using any suitable method of statistical analysis to establish a suitable baseline level using methods standard in the art for establishing such values. It will be appreciated by those of skill in the art that a baseline need not be established for each assay as the assay is performed but rather, a baseline can be established by referring to a form of stored information regarding a previously determined control level of gene expression. Such a form of stored information can include, for example, but is not limited to, a reference chart, listing or electronic file of population or individual data regarding "normal" (negative control)

or COPD-positive gene expression; a medical chart for the patient recording data from previous evaluations; or any other source of data regarding control gene expression that is useful for the patient to be diagnosed or evaluated.

Expression of the transcripts and/or proteins encoded by the genes of the invention is measured by any of a variety of known methods in the art. In general, the nucleic acid sequence of a nucleic acid molecule (*e.g.*, DNA or RNA) in a patient sample can be detected by any suitable method or technique of measuring or detecting gene sequence or expression. Such methods include, but are not limited to, polymerase chain reaction (PCR), reverse transcriptase-PCR (RT-PCR), *in situ* PCR, *in situ* hybridization, Southern blot, Northern blot, sequence analysis, microarray analysis, detection of a reporter gene, or other DNA/RNA hybridization platforms. For RNA expression, preferred methods include but are not limited to: extraction of cellular mRNA and Northern blotting using labeled probes that hybridize to transcripts encoding all or part of one or more of the genes of this invention; amplification of mRNA expressed from one or more of the genes of this invention using gene-specific primers, polymerase chain reaction (PCR), and reverse transcriptase-polymerase chain reaction (RT-PCR), followed by quantitative detection of the product by any of a variety of means; extraction of total RNA from the cells, which is then labeled and used to probe cDNAs or oligonucleotides encoding all or part of the genes of this invention, arrayed on any of a variety of surfaces; *in situ* hybridization; and detection of a reporter gene. The term "quantifying" or "quantitating" when used in the context of quantifying transcription levels of a gene can refer to absolute or to relative quantification. Absolute quantification may be accomplished by inclusion of known concentration(s) of one or more target nucleic acids and referencing the hybridization intensity of unknowns with the known target nucleic acids (*e.g.* through generation of a standard curve). Alternatively, relative quantification can be accomplished by comparison of hybridization signals between two or more genes, or between two or more treatments to quantify the changes in hybridization intensity and, by implication, transcription level.

The present invention includes isolated proteins encoded by the genes identified by the inventors as being differentially expressed in patients with COPD versus normal controls; that is, the proteins listed in Tables 2-5 and encoded by SEQ ID NOs: 1-____), isolated proteins encoded by a sequence complementary thereto, or polypeptides encoded by a fragment, homologue, or variant of genes represented in Tables 2-5. These proteins, peptides and polypeptides of the invention can be made using the genes or derived from the sequence information of the genes are also disclosed in the present invention. Functional forms of the proteins can be prepared, as purified preparations by using a cloned gene as described herein.

Alternatively, the proteins, peptides and polypeptides of the invention can be produced synthetically. Full length proteins or fragments corresponding to one or more particular motifs and/or domains or to arbitrary sizes, for example, at least about 5, 10, 25, 50, 75, or 100 amino acids in length are within the scope of the present invention.

5 Methods to measure protein expression levels of selected genes of this invention, include, but are not limited to: Western blot, immunoblot, enzyme-linked immunosorbant assay (ELISA), radioimmunoassay (RIA), immunoprecipitation, surface plasmon resonance, chemiluminescence, fluorescent polarization, phosphorescence, immunohistochemical analysis, matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry,
10 microcytometry, microarray, microscopy, fluorescence activated cell sorting (FACS), flow cytometry, and assays based on a property of the protein including but not limited to DNA binding, ligand binding, or interaction with other protein partners.

 Nucleic acid arrays are particularly useful for detecting the expression of the genes of the present invention. The production and application of high-density arrays in gene expression
15 monitoring have been disclosed previously in, for example, PCT Publication No. WO 97/10365; PCT Publication No. WO 92/10588; U.S. Patent No. 6,040,138; U.S. Patent No. 5,445,934; or PCT Publication No. WO 95/35505, all of which are incorporated herein by reference in their entireties. Also for examples of arrays, see Hacia *et al.* (1996) *Nature Genetics* 14:441-447; Lockhart *et al.* (1996) *Nature Biotechnol.* 14:1675-1680; and De Risi *et al.* (1996) *Nature Genetics* 14:457-460, each of which is incorporated by reference in its
20 entirety. In general, in an array, an oligonucleotide, a cDNA, or genomic DNA, that is a portion of a known gene, occupies a known location on a substrate. A nucleic acid target sample is hybridized with an array of such oligonucleotides and then the amount of target nucleic acids hybridized to each probe in the array is quantified. One preferred quantifying method is to use
25 confocal microscope and fluorescent labels. The *Affymetrix GeneChip™ Array* system (Affymetrix, Santa Clara, Calif.) and the *Atlas™ Human cDNA Expression Array* system are particularly suitable for quantifying the hybridization; however, it will be apparent to those of skill in the art that any similar systems or other effectively equivalent detection methods can also be used. In a particularly preferred embodiment, one can use the knowledge of the genes
30 described herein to design novel arrays of polynucleotides, cDNAs or genomic DNAs for screening methods described herein. Such novel pluralities of polynucleotides are contemplated to be a part of the present invention and are described in detail below.

 Suitable nucleic acid samples for screening on an array contain transcripts of interest or nucleic acids derived from the transcripts of interest (i.e., transcripts derived from the genes

associated with COPD of the present invention). As used herein, a nucleic acid derived from a transcript refers to a nucleic acid for whose synthesis the mRNA transcript or a subsequence thereof has ultimately served as a template. Thus, a cDNA reverse transcribed from a transcript, an RNA transcribed from that cDNA, a DNA amplified from the cDNA, an RNA
5 transcribed from the amplified DNA, etc., are all derived from the transcript and detection of such derived products is indicative of the presence and/or abundance of the original transcript in a sample. Thus, suitable samples include, but are not limited to, transcripts of the gene or genes, cDNA reverse transcribed from the transcript, cRNA transcribed from the cDNA, DNA amplified from the genes, RNA transcribed from amplified DNA, and the like. Preferably, such
10 a sample is a total RNA preparation of a biological sample (e.g., peripheral blood mononuclear cells or PBMCs). More preferably in some embodiments, such a nucleic acid sample is the total mRNA isolated from such a biological sample. Preferably, the nucleic acids for screening are obtained from a homogenate of cells (e.g., peripheral blood mononuclear cells or PBMCs).

In general, typical clinical samples include, but are not limited to, sputum, blood, blood
15 cells (e.g., peripheral blood mononuclear cells), tissue or fine needle biopsy samples, urine, peritoneal fluid, and pleural fluid, or cells therefrom. The present invention is primarily related to the detection of genes in peripheral blood mononuclear cells (PBMC or PBC).

In one embodiment, it is desirable to amplify the nucleic acid sample prior to hybridization. One of skill in the art will appreciate that whatever amplification method is used,
20 if a quantitative result is desired, care must be taken to use a method that maintains or controls for the relative frequencies of the amplified nucleic acids to achieve quantitative amplification. Methods of "quantitative" amplification are well known to those of skill in the art. For example, quantitative PCR involves simultaneously co-amplifying a known quantity of a control sequence using the same primers. This provides an internal standard that may be used to
25 calibrate the PCR reaction. The high-density array may then include probes specific to the internal standard for quantification of the amplified nucleic acid. Other suitable amplification methods include, but are not limited to polymerase chain reaction (PCR) Innis, et al., PCR Protocols. A guide to Methods and Application. Academic Press, Inc. San Diego, (1990)), ligase chain reaction (LCR) (see Wu and Wallace, Genomics, 4: 560 (1989), Landegren, et al.,
30 Science, 241: 1077 (1988) and Barringer, et al., Gene, 89: 117 (1990), transcription amplification (Kwoh, et al., Proc. Natl. Acad. Sci. USA, 86: 1173 (1989)), and self-sustained sequence replication (Guatelli, et al, Proc. Nat. Acad. Sci. USA, 87: 1874 (1990)).

Nucleic acid hybridization involves contacting a probe and target nucleic acid under conditions where the probe and its complementary target can form stable hybrid duplexes

through complementary base pairing. As used herein, hybridization conditions refer to standard hybridization conditions under which nucleic acid molecules are used to identify similar nucleic acid molecules. Such standard conditions are disclosed, for example, in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Labs Press, 1989. Sambrook et al., *ibid.*, is incorporated by reference herein in its entirety (see specifically, pages 9.31-9.62). In addition, formulae to calculate the appropriate hybridization and wash conditions to achieve hybridization permitting varying degrees of mismatch of nucleotides are disclosed, for example, in Meinkoth et al., 1984, *Anal. Biochem.* 138, 267-284; Meinkoth et al., *ibid.*, is incorporated by reference herein in its entirety. Nucleic acids that do not form hybrid duplexes are washed away from the hybridized nucleic acids and the hybridized nucleic acids can then be detected, typically through detection of an attached detectable label. It is generally recognized that nucleic acids are denatured by increasing the temperature or decreasing the salt concentration of the buffer containing the nucleic acids. Under low stringency conditions (e.g., low temperature and/or high salt) hybrid duplexes (e.g., DNA:DNA, RNA:RNA, or RNA:DNA) will form even where the annealed sequences are not perfectly complementary. Thus specificity of hybridization is reduced at lower stringency. Conversely, at higher stringency (e.g., higher temperature or lower salt) successful hybridization requires fewer mismatches.

High stringency hybridization and washing conditions, as referred to herein, refer to conditions which permit isolation of nucleic acid molecules having at least about 80% nucleic acid sequence identity with the nucleic acid molecule being used to probe in the hybridization reaction (i.e., conditions permitting about 20% or less mismatch of nucleotides). Very high stringency hybridization and washing conditions, as referred to herein, refer to conditions which permit isolation of nucleic acid molecules having at least about 90% nucleic acid sequence identity with the nucleic acid molecule being used to probe in the hybridization reaction (i.e., conditions permitting about 10% or less mismatch of nucleotides). As discussed above, one of skill in the art can use the formulae in Meinkoth et al., *ibid.* to calculate the appropriate hybridization and wash conditions to achieve these particular levels of nucleotide mismatch. Such conditions will vary, depending on whether DNA:RNA or DNA:DNA hybrids are being formed. Calculated melting temperatures for DNA:DNA hybrids are 10°C less than for DNA:RNA hybrids. In particular embodiments, stringent hybridization conditions for DNA:DNA hybrids include hybridization at an ionic strength of 6X SSC (0.9 M Na⁺) at a temperature of between about 20°C and about 35°C (lower stringency), more preferably, between about 28°C and about 40°C (more stringent), and even more preferably, between about 35°C and about 45°C (even more stringent), with appropriate wash conditions. In particular

embodiments, stringent hybridization conditions for DNA:RNA hybrids include hybridization at an ionic strength of 6X SSC (0.9 M Na⁺) at a temperature of between about 30°C and about 45°C, more preferably, between about 38°C and about 50°C, and even more preferably, between about 45°C and about 55°C, with similarly stringent wash conditions. These values are based on calculations of a melting temperature for molecules larger than about 100 nucleotides, 0% formamide and a G + C content of about 40%. Alternatively, T_m can be calculated empirically as set forth in Sambrook et al., *supra*, pages 9.31 to 9.62. In general, the wash conditions should be as stringent as possible, and should be appropriate for the chosen hybridization conditions. For example, hybridization conditions can include a combination of salt and temperature conditions that are approximately 20-25°C below the calculated T_m of a particular hybrid, and wash conditions typically include a combination of salt and temperature conditions that are approximately 12-20°C below the calculated T_m of the particular hybrid. One example of hybridization conditions suitable for use with DNA:DNA hybrids includes a 2-24 hour hybridization in 6X SSC (50% formamide) at about 42°C, followed by washing steps that include one or more washes at room temperature in about 2X SSC, followed by additional washes at higher temperatures and lower ionic strength (e.g., at least one wash as about 37°C in about 0.1X-0.5X SSC, followed by at least one wash at about 68°C in about 0.1X-0.5X SSC). Other hybridization conditions, and for example, those most useful with nucleic acid arrays, will be known to those of skill in the art.

The hybridized nucleic acids are detected by detecting one or more labels attached to the sample nucleic acids. The labels may be incorporated by any of a number of means well known to those of skill in the art. Detectable labels suitable for use in the present invention include any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful labels in the present invention include biotin for staining with labeled streptavidin conjugate, magnetic beads (e.g., DynabeadsTM), fluorescent dyes (e.g., fluorescein, texas red, rhodamine, green fluorescent protein, yellow fluorescent protein and the like), radiolabels (e.g., ³H, ¹²⁵I, ³⁵S, ¹⁴C, or ³²P), enzymes (e.g., horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic (e.g., polystyrene, polypropylene, latex, etc.) beads. Means of detecting such labels are well known to those of skill in the art. Thus, for example, radiolabels may be detected using photographic film or scintillation counters, fluorescent markers may be detected using a photodetector to detect emitted light. Enzymatic labels are typically detected by providing the enzyme with a substrate and detecting the reaction

product produced by the action of the enzyme on the substrate, and colorimetric labels are detected by simply visualizing the colored label.

The method of the present invention includes a step of comparing the results of detecting the expression of the one or more genes that are selectively regulated in patients with COPD as compared to a control (baseline normal or negative control) in order to determine whether there is any observed change or difference in expression of each gene in the patient as compared to the control. In one embodiment, a positive control (baseline COPD control) can also be used to assist in the confirmation of a diagnosis or prognosis. As discussed above, the present inventors have identified the expression profile of multiple genes that are differentially regulated in PBMCs of patients with COPD, as compared to a "normal" control (i.e., a patient that does not have or can not be detected to have COPD), including the manner in which the genes are regulated (i.e., up- or downregulated). Therefore, one can determine whether peripheral blood cells from a test patient have a gene expression profile that is statistically substantially similar to the profile of gene expression of a patient with COPD, or whether a profile of gene expression in the peripheral blood cells of the test patient is statistically more similar to the negative or normal, non-disease control. According to the present invention, an expression profile is substantially similar to a given profile of expression established for a group (e.g., COPD group, normal control group) if the expression profile of the gene or genes detected (including the identity of the gene, the manner in which expression is regulated, and/or the level of expression of the gene) is similar enough to the expected result so as to be statistically significant (i.e., with at least a 95% confidence level, or $p < 0.05$, and more preferably, with a confidence level of $p < 0.01$, and even more preferably, with a confidence level of $p < 0.005$, and even more preferably, with a confidence level of $p < 0.001$). Software programs and various statistical analysis techniques are available in the art that are capable of analyzing the expression of multiple genes and determining whether differences from a control are significant or not significant. For example, as discussed in the Examples, the gene expression measurements determined in patient samples were mean-centered and analyzed using the clustering, class comparison and class discovery functions of BRB ArrayTools and genes were selected that met the p value requirement ($p < 0.01$). In addition, statistical analysis methods are known in the art and described herein (see above and the Examples) that are preferably used to analyze the expression data generated for patient samples (e.g., independent and "leave-one-out" cross-validation and/or permutation testing).

By way of example, detection of the regulation of the expression of a gene in the "manner" associated with the established group, at a minimum, refers to the detection of the

regulation of a gene that has now been shown by the present inventors to be selectively regulated in PBMCs of patients having COPD, at least in the same direction (i.e., upregulation or downregulation) and preferably at a similar or comparable level, as compared to a normal or baseline control established for the expression of that gene. Preferably, a gene identified as
5 being upregulated or downregulated, as compared to a baseline control, is regulated in the same direction as the level of expression of the gene that is seen in established or confirmed patients with COPD as compared to a normal control. In other words, if “gene X” is upregulated in patients with COPD as compared to a normal control based on the inventors’ discovery presented herein, then one determines whether the expression of gene X is upregulated in
10 a patient test sample as compared to a normal control, or whether the expression of gene X is more similar to the level of expression of the normal control. In addition, a gene identified as being upregulated or downregulated, as compared to a baseline control, is regulated to at least about 10%, and more preferably at least 20%, and more preferably at least 25%, and more preferably at least 30%, and more preferably at least 35%, and more preferably at least 40%,
15 and more preferably at least 45%, and more preferably at least 50%, and preferably at least 55%, and more preferably at least 60%, and more preferably at least 65%, and more preferably at least 70%, and more preferably at least 75%, and more preferably at least 80%, and more preferably at least 85%, and more preferably at least 90%, and more preferably at least 95%, or even higher (e.g., above 100%) of the level of expression of the gene that is seen in established
20 or confirmed patients with COPD. Statistical significance should be at least $p < 0.05$, and more preferably, at least $p < 0.01$, and more preferably, $p < 0.005$, and even more preferably, $p < 0.001$. As discussed above, one of skill in the art can use software programs available in the art that use algorithms to analyze gene expression profiles and identify significant differences among samples and controls. In addition, one of skill in the art can apply various types of analyses as
25 discussed above (e.g., cross-validation and/or permutation testing) to validate the results of the methods described herein.

A profile of individual gene biomarkers identified in a method of the invention, including a matrix of two or more markers, can be generated by one or more of the methods described above. According to the present invention, a profile of the genes regulated in a
30 PBMC sample refers to a reporting of the expression level of a given gene that has been identified in any one or more of the tables presented herein, which, based on the knowledge of the regulation of the genes provided by the tables, includes a classification of the gene with regard to how the gene is regulated in PBMCs of a patient with COPD. For example, if a specific gene is identified as being expressed by a peripheral blood cell sample in a test patient,

the profile for the blood cell sample will include the reporting of the expression of this gene as compared to one or more baseline controls (e.g., a negative/normal and/or a positive/COPD control). Preferably, the profile includes data for more than one (e.g., at least two), and preferably several genes (e.g., at least five, six, seven, eight, nine, ten, or more genes), such that
5 a profile for the patient sample is created that can be compared to the control(s). The data can be reported as raw data, and/or statistically analyzed by any of a variety of methods, and/or combined with any other prognostic marker(s) for COPD, including any markers that are expressed in cells or tissues other than PBMCs and are useful for evaluating COPD in a patient. Prior to the present invention, one of skill in the art would not have known to screen patient
10 peripheral blood cells for the particular genes in the tables provided herein, and particularly for any combinations of these genes, and one of skill in the art would not have been able to classify these genes or combinations thereof on the basis of COPD versus normal.

It will be appreciated by those of skill in the art that differences between the expression of genes in PBMCs of patients with COPD and without COPD may be small or large. Some
15 small differences may be very reproducible and therefore are preferred for use in the diagnostic and prognostic methods of the invention. For other purposes, large differences may be desirable for ease of detection of the regulatory activity. It will therefore be appreciated that the exact boundary between a positive diagnosis and a negative diagnosis can shift, depending on the goal of the screening assay, the patient samples, the number of genes to be screened and the
20 baseline controls used. For some assays, a given patient may be sampled over time to detect the efficacy of a treatment, and so changes in gene expression from a disease state toward a normal state may be detected. In this case, the patient may still be positive for COPD as compared to a normal, disease-free control, but may show a shift toward the normal control gene expression profile if treatment is successful. In addition, the technique being used for detection, as well as
25 on the number of genes which are being tested, may impact how the assay is evaluated by those of skill in the art.

The profile of genes provided as a result of the screening of peripheral blood cells of a patient can be used by the patient or physician for decision-making regarding the usefulness of therapies for COPD in general. The profile can be used to estimate how the disease is likely to
30 respond and progress in any individual patient. Clinical trials can be developed to correlate the relationship between COPD regulated genes and the biological behavior of the diseased tissues, including in response to particular treatments for COPD.

In one aspect of this embodiment of the invention, the profiling of genes expressed by peripheral blood cells can be extended to other diseases, and particularly, to other pulmonary

diseases wherein diagnosis or prognosis of disease is difficult due to access to diseased tissue or difficulty distinguishing between subtypes of the disease based on conventional assays (e.g., histology). For example, as discussed above, using the guidance provided herein, it is within the ability of those of skill in the art to perform a *de novo* screening assay for the identification of genes regulated in peripheral blood cells in patients having a different disease, and particularly, a pulmonary disease, and to develop gene expression profiles for use in diagnostic and/or prognostic screening for these diseases. Moreover, one of skill in the art can use the techniques described herein to screen other gene arrays, including arrays of expressed tag sequences, to discover additional novel, genes that are regulated in the peripheral blood cells of patients with COPD. The extension of the gene profiles within COPD and also to other diseases will allow for the development of a variety of diagnostic assays in such diseases, as well as the identification of additional targets for therapeutic strategies.

It is to be understood that to perform the methods of the present invention, one of skill in the art can make use of any commercially available nucleotide or protein array, wherein hundreds or thousands of genes could be detected if desired. However, in the present method, one would use such an array to selectively screen only for the genes that are described as being useful for detection of COPD as disclosed herein, or to screen for such genes plus any other genes that would be useful as a predictor or analysis tool for COPD or for patients that have or may have COPD. In addition, the array can be designed to test for more than one disease condition in order to confirm or rule out other potential causes of a patient condition. For example, one may design an assay to screen for COPD as described herein, and also for pulmonary hypertension. In the specifically designed assays described herein, expression of the other non-informative genes in a large array can effectively be "ignored" or not screened. Alternatively, one of skill in the art can prepare nucleotide or protein arrays that are specifically designed to test for the expression of any combination of the genes of interest as described herein, alone or in combination with any other combination of genes that may be useful in evaluating a patient for COPD.

Another embodiment of the present invention relates to a plurality of polynucleotides for the detection of the expression of genes that are selectively regulated in peripheral blood cells of patients with COPD. The plurality of polynucleotides consists of, or consists essentially of, polynucleotide probes that are complementary to RNA transcripts, or nucleotides derived therefrom, of genes that have been identified herein as being selectively regulated in the peripheral blood cells of patients with COPD, and is therefore distinguished from previously known nucleic acid arrays and primer sets. The plurality of polynucleotides within the above-

limitation includes at least two or more polynucleotide probes (e.g., at least 2, 3, 4, 5, 6, and so on, in whole integer increments, up to all of the possible probes) that are complementary to RNA transcripts, or nucleotides derived therefrom, of genes identified by the present inventors. Such genes are selected from any of the genes listed in the tables provided herein. Multiple
5 probes can also be used to detect the same gene or to detect different splice variants of the same gene.

In one embodiment, it is contemplated that additional genes that are not regulated in the peripheral blood cells of patients with COPD, or that are not presently known to be regulated in the peripheral blood cells of patients with COPD, can be added to the set of genes to be
10 identified by the plurality of polynucleotides. Such genes would not be random genes, or large groups of unselected human genes, as are commercially available for detection now, but rather, would be specifically selected to complement the sets of genes identified by the present invention. For example, one of skill in the art may wish to add to the above-described plurality of polynucleotides one or more polynucleotides corresponding to (useful for identifying) genes
15 that are of relevance because they are expressed by a particular tissue of interest (e.g., pulmonary tissue), are associated with the particular disease (COPD) but not necessarily with peripheral blood cells, or are associated with a particular cell, tissue or body function. The development of additional pluralities of polynucleotides (and antibodies, as disclosed below), which include both the above-described plurality and such additional selected polynucleotides,
20 are explicitly contemplated by the present invention. In addition, using the techniques described herein, one of skill in the art may identify additional genes that are regulated in the peripheral blood cells of patients with COPD, and polynucleotides derived from such genes can be included in the plurality of polynucleotides described herein.

According to the present invention, a plurality of polynucleotides refers to at least 2, and
25 more preferably at least 3, and more preferably at least 4, and more preferably at least 5, and more preferably at least 6, and more preferably at least 7, and more preferably at least 8, and more preferably at least 9, and more preferably at least 10, and so on, in increments of one, up to any suitable number of polynucleotides, including polynucleotides representing all of the genes described herein (e.g., 106), 500, 1000, 10^4 , 10^5 , or at least 10^6 or more polynucleotides.

30 In accordance with the present invention, an isolated polynucleotide, or an isolated nucleic acid molecule, is a nucleic acid molecule that has been removed from its natural milieu (i.e., that has been subject to human manipulation), its natural milieu being the genome or chromosome in which the nucleic acid molecule is found in nature. As such, "isolated" does not necessarily reflect the extent to which the nucleic acid molecule has been purified, but

indicates that the molecule does not include an entire genome or an entire chromosome in which the nucleic acid molecule is found in nature. The polynucleotides useful in the plurality of polynucleotides of the present invention are typically a portion of a gene (sense or non-sense strand) of the present invention that is suitable for use as a hybridization probe or PCR primer for the identification of a full-length gene (or portion thereof) in a given sample (e.g., a peripheral blood cell sample). An isolated nucleic acid molecule can include a gene or a portion of a gene (e.g., the regulatory region or promoter), for example, to produce a reporter construct according to the present invention. An isolated nucleic acid molecule that includes a gene is not a fragment of a chromosome that includes such gene, but rather includes the coding region and regulatory regions associated with the gene, but no additional genes naturally found on the same chromosome. An isolated nucleic acid molecule can also include a specified nucleic acid sequence flanked by (i.e., at the 5' and/or the 3' end of the sequence) additional nucleic acids that do not normally flank the specified nucleic acid sequence in nature (i.e., heterologous sequences). Isolated nucleic acid molecule can include DNA, RNA (e.g., mRNA), or derivatives of either DNA or RNA (e.g., cDNA). Although the phrase "nucleic acid molecule" primarily refers to the physical nucleic acid molecule and the phrase "nucleic acid sequence" primarily refers to the sequence of nucleotides on the nucleic acid molecule, the two phrases can be used interchangeably, especially with respect to a nucleic acid molecule, or a nucleic acid sequence, being capable of encoding a protein. Preferably, an isolated nucleic acid molecule of the present invention is produced using recombinant DNA technology (e.g., polymerase chain reaction (PCR) amplification, cloning) or chemical synthesis.

The minimum size of a nucleic acid molecule or polynucleotide of the present invention is a size sufficient to encode a protein having a desired biological activity, sufficient to form a probe or oligonucleotide primer that is capable of forming a stable hybrid with the complementary sequence of a nucleic acid molecule encoding the natural protein (e.g., under moderate, high or very high stringency conditions), or to otherwise be used as a target in an assay or in any therapeutic method discussed herein. If the polynucleotide is an oligonucleotide probe or primer, the size of the polynucleotide can be dependent on nucleic acid composition and percent homology or identity between the nucleic acid molecule and a complementary sequence as well as upon hybridization conditions *per se* (e.g., temperature, salt concentration, and formamide concentration). The minimum size of a polynucleotide that is used as an oligonucleotide probe or primer is at least about 5 nucleotides in length, and preferably ranges from about 5 to about 50 or about 500 nucleotides or greater (1000, 2000, etc.), including any length in between, in whole number increments (i.e., 5, 6, 7, 8, 9, 10,...33, 34,...256,

257,...500...1000...), and more preferably from about 10 to about 40 nucleotides, and most preferably from about 15 to about 40 nucleotides in length. In one aspect, the oligonucleotide primer or probe is typically at least about 12 to about 15 nucleotides in length if the nucleic acid molecules are GC-rich and at least about 15 to about 18 bases in length if they are AT-rich.

5 There is no limit, other than a practical limit, on the maximal size of a nucleic acid molecule of the present invention, in that the nucleic acid molecule can include a portion of a protein-encoding sequence or a nucleic acid sequence encoding a full-length protein.

In one embodiment, the polynucleotide probes are conjugated to detectable markers. Detectable labels suitable for use in the present invention include any composition detectable by
10 spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful labels in the present invention include biotin for staining with labeled streptavidin conjugate, magnetic beads (e.g., Dynabeads.TM.), fluorescent dyes (e.g., fluorescein, texas red, rhodamine, green fluorescent protein, and the like), radiolabels (e.g., ³H, ¹²⁵I, ³⁵S, ¹⁴C, or ³²P), enzymes (e.g., horse radish peroxidase, alkaline phosphatase and others
15 commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic (e.g., polystyrene, polypropylene, latex, etc.) beads. Preferably, the polynucleotide probes are immobilized on a substrate.

In one embodiment, the polynucleotide probes are hybridizable array elements in a microarray or high density array. Nucleic acid arrays are well known in the art and are
20 described for use in comparing expression levels of particular genes of interest, for example, in U.S. Patent No. 6,177,248, which is incorporated herein by reference in its entirety. Nucleic acid arrays are suitable for quantifying a small variations in expression levels of a gene in the presence of a large population of heterogeneous nucleic acids. Knowing the identity of the genes set forth by the present invention, nucleic acid arrays can be fabricated either by *de novo*
25 synthesis on a substrate or by spotting or transporting nucleic acid sequences onto specific locations of substrate. Nucleic acids are purified and/or isolated from biological materials, such as a bacterial plasmid containing a cloned segment of sequence of interest. It is noted that all of the genes identified by the present invention have been previously sequenced, at least in part, such that oligonucleotides suitable for the identification of such nucleic acids can be produced.
30 The database accession number for each of the genes identified by the present inventors is provided in the tables of the invention. Suitable nucleic acids are also produced by amplification of template, such as by polymerase chain reaction or *in vitro* transcription.

One of skill in the art will appreciate that an enormous number of array designs are suitable for the practice of this invention. An array will typically include a number of probes

that specifically hybridize to the sequences of interest. In addition, in a preferred embodiment, the array will include one or more control probes. The high-density array chip includes "test probes". Test probes could be oligonucleotides having a minimum or maximum length as described above for other oligonucleotides. In another preferred embodiments, test probes are
5 double or single strand DNA sequences. DNA sequences are isolated or cloned from natural sources or amplified from natural sources using natural nucleic acids as templates, or produced synthetically. These probes have sequences complementary to particular subsequences of the genes whose expression they are designed to detect. Thus, the test probes are capable of specifically hybridizing to the target nucleic acid they are to detect.

10 Another embodiment of the present invention relates to a plurality of antibodies, or antigen binding fragments thereof, for the detection of the expression of genes regulated in peripheral blood cells in patients with COPD. The plurality of antibodies, or antigen binding fragments thereof, consists of antibodies, or antigen binding fragments thereof, that selectively bind to proteins encoded by genes that are regulated in peripheral blood cells in patients with
15 COPD, and that can be detected as protein products using antibodies. In addition, the plurality of antibodies, or antigen binding fragments thereof, comprises antibodies, or antigen binding fragments thereof, that selectively bind to proteins or portions thereof (peptides) encoded by any of the genes from the tables provided herein.

According to the present invention, a plurality of antibodies, or antigen binding
20 fragments thereof, refers to at least 2, and more preferably at least 3, and more preferably at least 4, and more preferably at least 5, and more preferably at least 6, and more preferably at least 7, and more preferably at least 8, and more preferably at least 9, and more preferably at least 10, and so on, in increments of one, up to any suitable number of antibodies, or antigen binding fragments thereof, including antibodies representing all of the genes described herein
25 (e.g., 246) or more, such as 500, or at least 1000 antibodies, or antigen binding fragments thereof.

According to the present invention, the phrase "selectively binds to" refers to the ability of an antibody, antigen binding fragment or binding partner (antigen binding peptide) to preferentially bind to specified proteins. More specifically, the phrase "selectively binds" refers
30 to the specific binding of one protein to another (e.g., an antibody, fragment thereof, or binding partner to an antigen), wherein the level of binding, as measured by any standard assay (e.g., an immunoassay), is statistically significantly higher than the background control for the assay. For example, when performing an immunoassay, controls typically include a reaction well/tube that contain antibody or antigen binding fragment alone (i.e., in the absence of antigen),

wherein an amount of reactivity (*e.g.*, non-specific binding to the well) by the antibody or antigen binding fragment thereof in the absence of the antigen is considered to be background. Binding can be measured using a variety of methods standard in the art including enzyme immunoassays (*e.g.*, ELISA), immunoblot assays, etc.).

5 Limited digestion of an immunoglobulin with a protease may produce two fragments. An antigen binding fragment is referred to as an Fab, an Fab', or an F(ab')₂ fragment. A fragment lacking the ability to bind to antigen is referred to as an Fc fragment. An Fab fragment comprises one arm of an immunoglobulin molecule containing a L chain (V_L + C_L domains) paired with the V_H region and a portion of the C_H region (CH1 domain). An Fab' 10 fragment corresponds to an Fab fragment with part of the hinge region attached to the CH1 domain. An F(ab')₂ fragment corresponds to two Fab' fragments that are normally covalently linked to each other through a di-sulfide bond, typically in the hinge regions.

Isolated antibodies of the present invention can include serum containing such antibodies, or antibodies that have been purified to varying degrees. Whole antibodies of the 15 present invention can be polyclonal or monoclonal. Alternatively, functional equivalents of whole antibodies, such as antigen binding fragments in which one or more antibody domains are truncated or absent (*e.g.*, Fv, Fab, Fab', or F(ab)₂ fragments), as well as genetically-engineered antibodies or antigen binding fragments thereof, including single chain antibodies or antibodies that can bind to more than one epitope (*e.g.*, bi-specific antibodies), or antibodies 20 that can bind to one or more different antigens (*e.g.*, bi- or multi-specific antibodies), may also be employed in the invention.

Generally, in the production of an antibody, a suitable experimental animal, such as, for example, but not limited to, a rabbit, a sheep, a hamster, a guinea pig, a mouse, a rat, or a chicken, is exposed to an antigen against which an antibody is desired. Typically, an animal is 25 immunized with an effective amount of antigen that is injected into the animal. An effective amount of antigen refers to an amount needed to induce antibody production by the animal. The animal's immune system is then allowed to respond over a pre-determined period of time. The immunization process can be repeated until the immune system is found to be producing antibodies to the antigen. In order to obtain polyclonal antibodies specific for the antigen, 30 serum is collected from the animal that contains the desired antibodies (or in the case of a chicken, antibody can be collected from the eggs). Such serum is useful as a reagent. Polyclonal antibodies can be further purified from the serum (or eggs) by, for example, treating the serum with ammonium sulfate.

Monoclonal antibodies may be produced according to the methodology of Kohler and Milstein (*Nature* 256:495-497, 1975). For example, B lymphocytes are recovered from the spleen (or any suitable tissue) of an immunized animal and then fused with myeloma cells to obtain a population of hybridoma cells capable of continual growth in suitable culture medium.

5 Hybridomas producing the desired antibody are selected by testing the ability of the antibody produced by the hybridoma to bind to the desired antigen.

Finally, any of the genes of this invention, or their RNA or protein products, can serve as targets for therapeutic strategies. For example, regulatory compounds that regulate (e.g., upregulate or downregulate) the expression and/or biological activity of a target gene or its
10 expression product (whether the product is intracellular, membrane or secreted) can be identified and/or designed using the information regarding the biomarker targets described herein. Alternatively, through the identification of particular genes that are highly regulated in patients with COPD, one can use such genes and their products to further investigate the molecular or biochemical mechanisms associated with the development and progression of
15 COPD and then design or establish assays to identify therapeutic compounds that affect the molecular or biochemical mechanism with the goal of providing a therapeutic benefit to the patient.

For example, one embodiment of the present invention relates to methods for identifying compounds that regulate the expression or activity of at least one of the biomarkers described
20 herein. Preferably, such compounds can be used to further study mechanisms associated with COPD or more preferably, serve as a therapeutic agent for use in the treatment or prevention of at least one symptom or aspect of COPD, or as a lead compound for the development of such a therapeutic agent. Once a biomarker has been identified as a target according to the present invention, an assay can be used for screening and selecting a chemical compound or a
25 biological compound having regulatory activity as a candidate reagent or therapeutic based on the ability of the compound to regulate the expression or activity of the target biomarker. Reference herein to regulating a target, can refer to one or both of regulating transcription of a target gene and regulating the translation and/or activity of its corresponding expression product. Such a compound can be referred to herein as therapeutic compound, in one
30 embodiment. For example, a cell line that naturally expresses the gene of interest or has been transfected with the gene (or suitable portions or derivatives thereof for assaying putative regulatory compounds) or other recombinant nucleic acid molecule encoding the protein of interest is incubated with various compounds, also referred to as candidate compounds, test compounds, or putative regulatory compounds. A regulation of the expression of the gene of

interest or regulation of the activities of its encoded product (*e.g.*, biological activity) may be used to identify a therapeutic compound. Therapeutic compounds identified in this manner can then be re-tested, if desired, in other assays to confirm their activities with regard to the target biomarker or a cellular or other activity related thereto.

5 In the method of the invention, the identification of compounds that increase the expression or activity of genes in any one or more of Tables 2-5, or the proteins encoded thereby, that are downregulated in peripheral blood cells of patients with COPD as compared to peripheral blood cells of normal controls, or the identification of compounds that decrease the expression or activity of genes in any one or more of Tables 2-5, or the proteins encoded
10 thereby, that are upregulated in peripheral blood cells of patients with COPD as compared to peripheral blood cells of normal controls, are predicted to be useful as therapeutic reagents or lead compounds therefore in the prevention and treatment of COPD.

For example one embodiment of the present invention relates to a method of using the differentially expressed genes described herein or the proteins encoded thereby (*i.e.*, the
15 biomarkers of the invention) as a target to identify a regulatory compound for regulation of a biological function associated with that gene or protein. Such a method can include the steps of: (a) contacting a test compound with a cell that expresses the target biomarker or a useful portion thereof (*i.e.*, useful being any portion of a gene, transcript or protein that can be used to identify a compound as discussed herein); and (b) identifying compounds that regulate the
20 expression or activity of the gene or protein.

In general, the biological activity or biological action of a protein refers to any function(s) exhibited or performed by the protein that is ascribed to the naturally occurring form of the protein as measured or observed *in vivo* (*i.e.*, in the natural physiological environment of the protein) or *in vitro* (*i.e.*, under laboratory conditions). Modifications, activities or
25 interactions which result in a decrease in protein expression or a decrease in the activity of the protein, can be referred to as inactivation (complete or partial), down-regulation, reduced action, or decreased action or activity of a protein. Similarly, modifications, activities or interactions which result in an increase in protein expression or an increase in the activity of the protein, can be referred to as amplification, overproduction, activation, enhancement, up-
30 regulation or increased action of a protein. The biological activity of a protein according to the invention can be measured or evaluated using any assay for the biological activity of the protein as known in the art. Such assays can include, but are not limited to, binding assays, assays to determine internalization of the protein and/or associated proteins, enzyme assays, cell signal transduction assays (*e.g.*, phosphorylation assays), and/or assays for determining downstream

cellular events that result from activation or binding of the cell surface protein (*e.g.*, expression of downstream genes, production of various biological mediators, etc.).

According to the present invention, a biologically active fragment or homologue of a gene, nucleic acid transcript or derivative thereof, or protein maintains the ability to be useful in a method of the present invention. Therefore, the biologically active fragment or homologue maintains the ability to be used to identify regulators (*e.g.*, inhibitors) of the native gene or protein when, for example, the biologically active fragment or homologue is expressed by a cell or used in another assay format. Therefore, the biologically active fragment or homologue has a structure that is sufficiently similar to the structure of the native gene or protein that a regulatory compound can be identified by its ability to bind to and/or regulate the expression or activity of the fragment or homologue in a manner consistent with the regulation of the native gene or protein.

Compounds to be screened in the methods of the invention include known organic compounds such as antibodies, products of peptide libraries, and products of chemical combinatorial libraries. Compounds may also be identified using rational drug design relying on the structure of the product of a gene. Such methods are known to those of skill in the art and involve the use of three-dimensional imaging software programs. For example, various methods of drug design, useful to design or select mimetics or other therapeutic compounds useful in the present invention are disclosed in Maulik et al., 1997, *Molecular Biotechnology: Therapeutic Applications and Strategies*, Wiley-Liss, Inc., which is incorporated herein by reference in its entirety.

As used herein, a mimetic refers to any peptide or non-peptide compound that is able to mimic the biological action of a naturally occurring peptide, often because the mimetic has a basic structure that mimics the basic structure of the naturally occurring peptide and/or has the salient biological properties of the naturally occurring peptide. Mimetics can include, but are not limited to: peptides that have substantial modifications from the prototype such as no side chain similarity with the naturally occurring peptide (such modifications, for example, may decrease its susceptibility to degradation); anti-idiotypic and/or catalytic antibodies, or fragments thereof; non-proteinaceous portions of an isolated protein (*e.g.*, carbohydrate structures); or synthetic or natural organic molecules, including nucleic acids and drugs identified through combinatorial chemistry, for example. Such mimetics can be designed, selected and/or otherwise identified using a variety of methods known in the art.

A mimetic can be obtained, for example, from molecular diversity strategies (a combination of related strategies allowing the rapid construction of large, chemically diverse

molecule libraries), libraries of natural or synthetic compounds, in particular from chemical or combinatorial libraries (*i.e.*, libraries of compounds that differ in sequence or size but that have the similar building blocks) or by rational, directed or random drug design. *See* for example, Maulik et al., *supra*.

5 In a molecular diversity strategy, large compound libraries are synthesized, for example, from peptides, oligonucleotides, carbohydrates and/or synthetic organic molecules, using biological, enzymatic and/or chemical approaches. The critical parameters in developing a molecular diversity strategy include subunit diversity, molecular size, and library diversity. The general goal of screening such libraries is to utilize sequential application of combinatorial
10 selection to obtain high-affinity ligands for a desired target, and then to optimize the lead molecules by either random or directed design strategies. Methods of molecular diversity are described in detail in Maulik, et al., *ibid*.

Maulik et al. also disclose, for example, methods of directed design, in which the user directs the process of creating novel molecules from a fragment library of appropriately selected
15 fragments; random design, in which the user uses a genetic or other algorithm to randomly mutate fragments and their combinations while simultaneously applying a selection criterion to evaluate the fitness of candidate ligands; and a grid-based approach in which the user calculates the interaction energy between three dimensional receptor structures and small fragment probes, followed by linking together of favorable probe sites.

20 As used herein, the term "test compound", "putative inhibitory compound" or "putative regulatory compound" refers to compounds having an unknown or previously unappreciated regulatory activity in a particular process. As such, the term "identify" with regard to methods to identify compounds is intended to include all compounds, the usefulness of which as a regulatory compound for the purposes of regulating the expression or activity of a target
25 biomarker or otherwise regulating some activity that may be useful in the study or treatment of COPD is determined by a method of the present invention.

In one embodiment of the invention, regulatory compounds are identified by exposing a target gene to a test compound; measuring the expression of a target; and selecting a compound that regulates (up or down) the expression of the target. For example, the putative regulatory
30 compound can be exposed to a cell that expresses the target gene (endogenously or recombinantly). A preferred cell to use in an assay includes a mammalian cell that either naturally expresses the target gene or has been transformed with a recombinant form of the target gene, such as a recombinant nucleic acid molecule comprising a nucleic acid sequence

encoding the target protein or a useful fragment thereof. Methods to determine expression levels of a gene are well known in the art.

The conditions under which a cell, cell lysate, nucleic acid molecule or protein of the present invention is exposed to or contacted with a putative regulatory compound, such as by mixing, are any suitable culture or assay conditions. In the case of a cell-based assay, the conditions include an effective medium in which the cell can be cultured or in which the cell lysate can be evaluated in the presence and absence of a putative regulatory compound. Cells of the present invention can be cultured in a variety of containers including, but not limited to, tissue culture flasks, test tubes, microtiter dishes, and petri plates. Culturing is carried out at a temperature, pH and carbon dioxide content appropriate for the cell. Such culturing conditions are also within the skill in the art. Cells are contacted with a putative regulatory compound under conditions which take into account the number of cells per container contacted, the concentration of putative regulatory compound(s) administered to a cell, the incubation time of the putative regulatory compound with the cell, and the concentration of compound administered to a cell. Determination of effective protocols can be accomplished by those skilled in the art based on variables such as the size of the container, the volume of liquid in the container, conditions known to be suitable for the culture of the particular cell type used in the assay, and the chemical composition of the putative regulatory compound (*i.e.*, size, charge etc.) being tested. A preferred amount of putative regulatory compound(s) can comprise between about 1 nM to about 10 mM of putative regulatory compound(s) per well of a 96-well plate.

To detect expression of a target refers to the act of actively determining whether a target is expressed or not. This can include determining whether the target expression is upregulated as compared to a control, downregulated as compared to a control, or unchanged as compared to a control. Therefore, the step of detecting expression does not require that expression of the target actually is upregulated or downregulated, but rather, can also include detecting that the expression of the target has not changed (*i.e.*, detecting no expression of the target or no change in expression of the target). Expression of transcripts and/or proteins is measured by any of a variety of known methods in the art, and such methods have been discussed previously herein. Similarly, measurement of translation of a protein includes any suitable method for detecting and/or measuring proteins from a cell or cell extract, and such methods have been described previously herein.

Designing a compound for testing in a method of the present invention can include creating a new chemical compound or searching databases of libraries of known compounds (*e.g.*, a compound listed in a computational screening database containing three dimensional

structures of known compounds). Designing can also be performed by simulating chemical compounds having substitute moieties at certain structural features. The step of designing can include selecting a chemical compound based on a known function of the compound. A preferred step of designing comprises computational screening of one or more databases of compounds in which the three dimensional structure of the compound is known and is interacted (*e.g.*, docked, aligned, matched, interfaced) with the three dimensional structure of a target by computer (*e.g.* as described by Humblet and Dunbar, *Animal Reports in Medicinal Chemistry*, vol. 28, pp. 275-283, 1993, M Venuti, ed., Academic Press). Methods to synthesize suitable chemical compounds are known to those of skill in the art and depend upon the structure of the chemical being synthesized. Methods to evaluate the bioactivity of the synthesized compound depend upon the bioactivity of the compound (*e.g.*, inhibitory or stimulatory).

Candidate compounds identified or designed by the above-described methods can be synthesized using techniques known in the art, and depending on the type of compound. Synthesis techniques for the production of non-protein compounds, including organic and inorganic compounds are well known in the art. For example, for smaller peptides, chemical synthesis methods are preferred. For example, such methods include well known chemical procedures, such as solution or solid-phase peptide synthesis, or semi-synthesis in solution beginning with protein fragments coupled through conventional solution methods. Such methods are well known in the art and may be found in general texts and articles in the area such as: Merrifield, 1997, *Methods Enzymol.* 289:3-13; Wade et al., 1993, *Australas Biotechnol.* 3(6):332-336; Wong et al., 1991, *Experientia* 47(11-12):1123-1129; Carey et al., 1991, *Ciba Found Symp.* 158:187-203; Plaue et al., 1990, *Biologicals* 18(3):147-157; Bodanszky, 1985, *Int. J. Pept. Protein Res.* 25(5):449-474; or H. Dugas and C. Penney, BIOORGANIC CHEMISTRY, (1981) at pages 54-92, all of which are incorporated herein by reference in their entirety. For example, peptides may be synthesized by solid-phase methodology utilizing a commercially available peptide synthesizer and synthesis cycles supplied by the manufacturer. One skilled in the art recognizes that the solid phase synthesis could also be accomplished using the Fmoc strategy and a TFA/scavenger cleavage mixture. A compound that is a protein or peptide can also be produced using recombinant DNA technology and methods standard in the art, particularly if larger quantities of a protein are desired.

In another embodiment of the invention, putative regulatory compounds are identified by exposing a target to a candidate compound; measuring the binding of the candidate compound to the target; and selecting a compound that binds to the target at a desired

concentration, affinity, or avidity. In a preferred embodiment, the assay is performed under conditions conducive to promoting the interaction or binding of the compound to the target. One of skill in the art can determine such conditions based on the target and the compound being used in the assay. In one embodiment, a BIAcore machine can be used to determine the
5 binding constant of a complex between the target protein (a protein encoded by the target gene) and a natural ligand in the presence and absence of the candidate compound. For example, the target protein or a ligand binding fragment thereof can be immobilized on a substrate. A natural or synthetic ligand is contacted with the substrate to form a complex. The dissociation constant for the complex can be determined by monitoring changes in the refractive index with respect to
10 time as buffer is passed over the chip (O'Shannessy et al. Anal. Biochem. 212:457-468 (1993); Schuster et al., Nature 365:343-347 (1993)). Contacting a candidate compound at various concentrations with the complex and monitoring the response function (*e.g.*, the change in the refractive index with respect to time) allows the complex dissociation constant to be determined in the presence of the test compound and indicates whether the candidate compound is either an
15 inhibitor or an agonist of the complex. Alternatively, the candidate compound can be contacted with the immobilized target protein at the same time as the ligand to see if the candidate compound inhibits or stabilizes the binding of the ligand to the target protein.

Other suitable assays for measuring the binding of a candidate compound to a target protein or for measuring the ability of a candidate compound to affect the binding of the target
20 protein to another protein or molecule include, but are not limited to, Western blot, immunoblot, enzyme-linked immunosorbant assay (ELISA), radioimmunoassay (RIA), immunoprecipitation, surface plasmon resonance, chemiluminescence, fluorescent polarization, phosphorescence, immunohistochemical analysis, matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry, microcytometry, microarray, microscopy, fluorescence
25 activated cell sorting (FACS), and flow cytometry. Other assays include those that are suitable for monitoring the effects of protein binding, including, but not limited to, cell-based assays such as: cytokine secretion assays, or intracellular signal transduction assays that determine, for example, protein or lipid phosphorylation, mediator release or intracellular Ca⁺⁺ mobilization.

In yet another embodiment, putative regulatory compounds are identified by exposing a
30 target protein of the present invention (or a cell expressing the protein naturally or recombinantly) to a candidate compound and measuring the ability of the compound to inhibit or enhance a biological activity of the protein. In one embodiment, the biological activity of a protein encoded by the target gene is measured by measuring the amount of product generated in a biochemical reaction mediated by the protein encoded by the target gene. In still another

embodiment, the activity of the protein encoded by the target gene is measured by measuring the amount of substrate generated in a biochemical reaction mediated by the protein encoded by the target gene. In another embodiment, a biological activity is measured by measuring a specific event in a cell-based assay, such as release or secretion of a biological mediator or compound that is regulated by the activity of the target protein, measuring intracellular signal transduction assays that determine, for example, protein or lipid phosphorylation, mediator release or intracellular Ca^{++} mobilization. Preferably, the activity of the protein is measured in the presence and absence of the candidate compound, or in the presence of another suitable control compound.

In one embodiment of the invention, when the protein encoded by a target gene is an enzyme, a therapeutic compound is identified by exposing the enzyme encoded by a target gene to a test compound; measuring the activity of the enzyme encoded by the target gene in the presence and absence of the compound; and selecting a compound that down-regulates or inhibits the activity of the enzyme encoded by the target gene. Methods to measure enzymatic activity are well known to those skilled in the art and are selected based on the identity of the enzyme being tested. For example, if the enzyme is a kinase, phosphorylation assays can be used.

Preferably, methods used to identify therapeutic compounds are customized for each target gene or product. For example, if the target product is an enzyme, then the enzyme will be expressed in cell culture and purified. The enzyme will then be screened *in vitro* against therapeutic compounds to look for inhibition of that enzymatic activity. If the target is a non-catalytic protein, then it will also be expressed and purified. Therapeutic compounds will then be tested for their ability to regulate, for example, the binding of a site-specific antibody or a target-specific ligand to the target product.

In a preferred embodiment, therapeutic compounds that bind to target products are identified, then those compounds can be further tested in biological assays that test for other desirable characteristics and activities, such as utility as a reagent for the study of COPD or utility as a therapeutic compound for the prevention or treatment of COPD.

If a suitable therapeutic compound is identified using the methods and genes of the present invention, a composition can be formulated. A composition, and particularly a therapeutic composition, of the present invention generally includes the therapeutic compound and a carrier, and preferably, a pharmaceutically acceptable carrier. According to the present invention, a "pharmaceutically acceptable carrier" includes pharmaceutically acceptable excipients and/or pharmaceutically acceptable delivery vehicles, which are suitable for use in

administration of the composition to a suitable *in vitro*, *ex vivo* or *in vivo* site. A suitable *in vitro*, *in vivo* or *ex vivo* site is preferably a pulmonary tissue or a cell that is associated with or travels to a pulmonary tissue. Preferred pharmaceutically acceptable carriers are capable of maintaining a compound, a protein, a peptide, nucleic acid molecule or mimetic (drug) in a form that, upon arrival of the compound, protein, peptide, nucleic acid molecule or mimetic at the target site in a culture (in the case of an *in vitro* or *ex vivo* protocol) or in patient (*in vivo*), the compound, protein, peptide, nucleic acid molecule or mimetic is capable of providing the desired effect at the target site.

Suitable excipients of the present invention include excipients or formularies that transport or help transport, but do not specifically target a composition to a cell (also referred to herein as non-targeting carriers). Examples of pharmaceutically acceptable excipients include, but are not limited to water, phosphate buffered saline, Ringer's solution, dextrose solution, serum-containing solutions, Hank's solution, other aqueous physiologically balanced solutions, oils, esters and glycols. Aqueous carriers can contain suitable auxiliary substances required to approximate the physiological conditions of the recipient, for example, by enhancing chemical stability and isotonicity.

One type of pharmaceutically acceptable carrier includes a controlled release formulation that is capable of slowly releasing a composition of the present invention into a patient or culture. As used herein, a controlled release formulation comprises a therapeutic compound in a controlled release vehicle. Suitable controlled release vehicles include, but are not limited to, biocompatible polymers, other polymeric matrices, capsules, microcapsules, microparticles, bolus preparations, osmotic pumps, diffusion devices, liposomes, lipospheres, and transdermal delivery systems. Other carriers include liquids that, upon administration to a patient, form a solid or a gel *in situ*. Preferred carriers are also biodegradable (i.e., bioerodible). When the compound is a recombinant nucleic acid molecule, suitable delivery vehicles include, but are not limited to liposomes, viral vectors or other delivery vehicles, including ribozymes. Natural lipid-containing delivery vehicles include cells and cellular membranes. Artificial lipid-containing delivery vehicles include liposomes and micelles. A delivery vehicle of the present invention can be modified to target to a particular site in a patient, thereby targeting and making use of a therapeutic compound at that site. Suitable modifications include manipulating the chemical formula of the lipid portion of the delivery vehicle and/or introducing into the vehicle a targeting agent capable of specifically targeting a delivery vehicle to a preferred site, for example, a preferred cell type. Other suitable delivery vehicles include gold particles, poly-L-lysine/DNA-molecular conjugates, and artificial chromosomes.

A compound or composition can be delivered to a cell culture or patient by any suitable method. Selection of such a method will vary with the type of compound being administered or delivered (i.e., compound, protein, peptide, nucleic acid molecule, or mimetic), the mode of delivery (i.e., *in vitro*, *in vivo*, *ex vivo*) and the goal to be achieved by administration/delivery of the compound or composition. According to the present invention, an effective administration protocol (i.e., administering a composition in an effective manner) comprises suitable dose parameters and modes of administration that result in delivery of a composition to a desired site (i.e., to a desired cell) and/or in the desired regulatory event.

Administration routes include *in vivo*, *in vitro* and *ex vivo* routes. *In vivo* routes include, but are not limited to, oral, nasal, intratracheal injection, inhaled, transdermal, rectal, and parenteral routes. Preferred parenteral routes can include, but are not limited to, subcutaneous, intradermal, intravenous, intramuscular and intraperitoneal routes. Intravenous, intraperitoneal, intradermal, subcutaneous and intramuscular administrations can be performed using methods standard in the art. Aerosol (inhalation) delivery can also be performed using methods standard in the art (see, for example, Stribling et al., *Proc. Natl. Acad. Sci. USA* 189:11277-11281, 1992, which is incorporated herein by reference in its entirety). Oral delivery can be performed by complexing a therapeutic composition of the present invention to a carrier capable of withstanding degradation by digestive enzymes in the gut of an animal. Examples of such carriers, include plastic capsules or tablets, such as those known in the art. Direct injection techniques include, for example, injecting the composition directly into a site. *Ex vivo* refers to performing part of the regulatory step outside of the patient, such as by transfecting a population of cells removed from a patient with a recombinant molecule comprising a nucleic acid sequence encoding a protein according to the present invention under conditions such that the recombinant molecule is subsequently expressed by the transfected cell, and returning the transfected cells to the patient. *In vitro* and *ex vivo* routes of administration of a composition to a culture of host cells can be accomplished by a method including, but not limited to, transfection, transformation, electroporation, microinjection, lipofection, adsorption, protoplast fusion, use of protein carrying agents, use of ion carrying agents, use of detergents for cell permeabilization, and simply mixing (e.g., combining) a compound in culture with a target cell.

In the method of the present invention, a therapeutic compound, as well as compositions comprising such compounds, can be administered to any organism, and particularly, to any member of the Vertebrate class, Mammalia, including, without limitation, primates, rodents, livestock and domestic pets. Livestock include mammals to be consumed or that produce

useful products (e.g., sheep for wool production). Preferred mammals to protect include humans.

Typically, it is desirable to obtain a therapeutic benefit in a patient. A therapeutic benefit is not necessarily a cure for a particular disease or condition, but rather, preferably encompasses a result which can include alleviation of the disease or condition, elimination of the disease or condition, reduction of a symptom associated with the disease or condition, prevention or alleviation of a secondary disease or condition resulting from the occurrence of a primary disease or condition, and/or prevention of the disease or condition. As used herein, the phrase "protected from a disease" refers to reducing the symptoms of the disease, reducing the occurrence of the disease, and/or reducing the severity of the disease. Protecting a patient can refer to the ability of a composition of the present invention, when administered to a patient, to prevent a disease from occurring and/or to cure or to alleviate disease symptoms, signs or causes. As such, to protect a patient from a disease includes both preventing disease occurrence (prophylactic treatment) and treating a patient that has a disease (therapeutic treatment) to reduce the symptoms of the disease. A beneficial effect can easily be assessed by one of ordinary skill in the art and/or by a trained clinician who is treating the patient. The term, "disease" refers to any deviation from the normal health of a mammal and includes a state when disease symptoms are present, as well as conditions in which a deviation (e.g., infection, gene mutation, genetic defect, etc.) has occurred, but symptoms are not yet manifested.

20

The following examples are provided for the purpose of illustration and are not intended to limit the scope of the present invention.

Examples

25 Example 1

The following example demonstrates the utility of peripheral blood mononuclear cells as surrogate markers for COPD.

The key initial discovery of the present inventors' studies was the demonstration of the utility of PBMC gene expression as a surrogate marker for COPD. A recent study has shown that the gene expression profile of PBMC compared among normal individuals is remarkably homogeneous, and distinct from the expression profiles of diseased individuals⁵⁴. In the first experiments, the present inventors collected PBMC data collected on five former smokers with advanced COPD (Fig. 3). This group was used as a comparative group in a study of pulmonary arterial hypertension (PAH) versus normal controls (non-PAH samples) to validate distinct

immunologic profiles within different disease processes. The PAH study is described in detail in Bull *et al.*¹³, and is incorporated herein by reference in its entirety, although the COPD data presented herein is described for the first time herein.

5 The Bull *et al.* study was designed to emphasize the importance of measure validation, which is again used to evaluate the gene expression in the experiments of the present invention. Specifically, due to the high measurement-to-sample ratio inherent in microarray studies, steps must be taken to avoid “over-fitting” any predictive model⁵⁵. For this reason, both independent and “leave-one-out” cross-validation of the predictor was employed. In addition permutation testing was used to assess the significance of the cross validation error rate. This procedure
10 was performed by randomly permuting class labels among the gene expression measurements (for example normal vs. COPD) considering 2000 permutations of the class labels. The proportion of the permuted data sets that have a classification error rate less than or equal to the unpermuted misclassification rate serve as the achieved significance level in a test against the null hypothesis - that there is no difference in gene expression profiles between the two classes.
15 Finally, validation was performed using an independent measure of gene expression magnitude, and the differences in gene expression were demonstrated again using this independent measure of mRNA abundance. These measurements were extended in the PAH study with a larger cohort with q-PCR analysis of two genes. The power of the microarray analysis to determine biomarkers is made evident by the fact that the larger cohort also demonstrated the differences
20 in gene expression profiles predicted by the smaller subset of diseased and normal patients.

As shown in Fig. 3, the initial comparisons with PBMCs from age and gender matched non-smokers that have COPD supports distinct expression profiles between the groups. In particular, immunologic profiles including cytokine and chemokine mediators, major histocompatibility antigen markers, G-protein coupled receptors, and markers of oxidative
25 stress are differentially expressed (Table 2 shows the entire list of 246 transcripts). The present inventors' prior experience using this technique in other disease models has consistently led to statistical confirmation of similar preliminary results when performed with larger samples. Therefore, larger samplings are predicted to support these conclusions.

The initial PBMC expression profile data in COPD described above supports several
30 important points. First, the inventors have demonstrated the technical requirements for successfully isolating mononuclear cells from peripheral blood and extracting high quality RNA of appropriate quantity to generate microarray gene expression profiles. Second, PBMC gene expression from individuals with COPD has a distinct immunologic phenotype from normal controls and patients with other chronic lung diseases (PAH). Lastly, the differentially

expressed transcripts in the PBMC of patients with COPD preliminary support for the inventors' hypothesis that immunoregulatory differences are present between tobacco exposed individuals with and without COPD.

Example 2

5 The following example describes a larger study demonstrating the utility of peripheral blood mononuclear cells as surrogate markers for COPD.

Experimental Design

10 The outline of the study is as follows. Four groups (n=10) of tobacco exposed individuals are studied: (1) current smokers with at least Stage I GOLD classification COPD⁵⁶, (2) former smokers with at least Stage I GOLD classification COPD, (3) current smokers without physiologic evidence of COPD matched to group 1 for quantitative smoking history, age, gender, and ethnicity, and (4) former smokers without physiologic evidence of COPD matched to group 2 for quantitative smoking history, age, gender, and ethnicity. For groups (1) and (2), Exclusion Criteria are: history of underlying lung disease other than COPD (e.g. 15 asthma) or systemic disease with known pulmonary manifestations (rheumatoid arthritis; systemic lupus erythematosus), etc.); known history of genetic predisposition to COPD (alpha-one antitrypsin, cystic fibrosis, etc.) or potential occupational exposure (mining, metal work, etc.); use of immunosuppressive medication including oral or inhaled corticosteroids; chronic oxygen therapy; and history of malignancy. Additional inclusion criteria for groups (3) and (4) 20 includes: no evidence of airflow limitation by spirometry; FEV1 > 80% predicted; and FEV1/FVC ratio > 70. Exclusion criteria for groups (3) and (4) includes: history of underlying lung disease (e.g. asthma, chronic bronchitis, etc.) or systemic disease with known pulmonary manifestations (e.g. SLE, RA, etc.); use of immunosuppressive medications; and history of malignancy. In addition to array analysis on all individuals, RNA isolated from PBMC 25 separation will be stored for directed (quantitative reverse transcriptase PCR) validation analysis of observed differentially expressed genes. At time of initial blood draw, an additional peripheral PBMC isolation tube will be collected and cryopreserved for future validation studies of differentially expressed cell marker transcripts (Fig. 4).

Sample size and power

30 Using the data presented in Example 1 on differential gene expression, the inventors have determined that a sample size of 10 per group will allow the identification of large differences between any two groups. Table 2 shows minimally detectable mean differences (Mean2-Mean1) on the log₂ intensity scale using an α -level of 0.001 (two-sided) and power of 80%⁵⁷.

Table 2: sample size and power.

STD Value	Power	N1	N2	Alpha	Mean N1	Mean N2	Fold change	S1	S2	Effect size
Median	0.80	10	10	0.001	0.0	0.653	1.6	0.30	0.30	1.09
75 th percentile	0.80	10	10	0.001	0.0	0.980	2.0	0.45	0.45	1.09
90 th percentile	0.80	10	10	0.001	0.0	1.524	2.9	0.70	0.70	1.09

The estimated standard deviations (S1 and S2) were obtained from the COPD and normal distributions of standard deviations from the 15,022 genes that passed the filtering criteria applied in BRB ArrayTools. Values are shown for the median, 75th and 90th percentiles of standard deviation as suggested by Yang and Speed⁵⁸. It is assumed for these calculations that an unmatched analysis will be performed (or that the matching will not induce correlation between COPD and non-COPD individuals), which gives a conservative estimate of effect size. As the true proportion of differentially expressed genes in the population varies from 0.005, 10 0.05 to 0.20, the expected false discovery rate will vary from 0.20, 0.02, to 0.005⁵⁹.

Isolation of PBMCs and total RNA

The mononuclear cells are obtained from patient blood in an identical manner to that employed for the PAH study discussed in Example 1, with slight modifications to improve reproducibility. Specifically, eight milliliters of peripheral blood is collected into BD-CPT tubes⁵⁴ and processed following the manufacturer's instructions. In recent testing, the inventors have determined that PBMCs isolated in this fashion contain less than 5% granulocytes. Total RNA from samples which are selected for array analysis are isolated using standard methods, quantified by spectrophotometry, and the absence of degradation assured using the Agilent Bioanalyzer.

Target labeling and array hybridization

Biotinylated cRNA for array hybridization is generated from total RNA using the methods previously developed by the inventors and others. Briefly, total RNA (2-5 μ g) is converted to double-stranded cDNA using a standard oligo-dT-T7 primer, followed by in vitro transcription by T7 with the incorporation of biotin-nucleotide triphosphates. Labeled cRNA is fragmented, added to a hybridization buffer, and applied to the microarray. Subsequent to the hybridization the unbound cRNA is washed away, and the bound probe is stained with Streptavidin Phycoerythrin. The array is scanned, and the quantity of hybridization is inferred from the intensity of fluorescence at each feature of the array.

The most current Affymetrix® arrays for human gene expression are used: e.g., the 30 Affymetrix Hu-133 Plus 2.0 Genechip™. This fifth-generation microarray measures

approximately 47,000 transcripts, including essentially all of the well-characterized genes in the human genome (*e.g.*, ~38,500 genes). Analysis of this microarray includes the use of a high-resolution scanner that is demanded for arrays with this increased density (1,300,000 individual array features).

5 *Microarray data analysis*

The first step of the data flow is the conversion of image data into tabular data. The inventors use the statistical algorithms implemented in *Affymetrix GeneChip™* Operating System (GCOS) (Affymetrix, Santa Clara, Calif.) for this task. While a number of alternatives to GCOS exist (such as d-chip, RMA, and PerfectMatch), the inventors' experience indicates that while these tools may provide advantages in the direct comparison of paired samples, they provide no advantage in class comparison and class discovery applications. Internal measurements of chip and sample quality (brightness scaling factor, noise, % present calls, and control gene 3'/5' ratios) are collected in GCOS, and only chips that meet the inventors' quality threshold are included for analysis.

15 The bulk of data analysis is conducted with the software package BRB ArrayTools v3.0.2E developed by Dr Richard Simon and Amy Peng Lam at the National Cancer Institute. This software emphasizes conservative, defensible analysis of array data, with a high level of validation through the use of permutation testing and leave-one-out cross-validation with removal of the left-out data set prior to gene selection. This approach leads to the association of reliability estimates with each discovery.

20 The analysis of the array data is similar to that outlined for the pulmonary hypertension analysis in Example 1 above. Briefly, the data is normalized (mean centered) and examined in an unsupervised fashion (see description of study design above). The goal of this analysis is to reduce bias in the discovery of meaningful groups in the patient population. Subsequent supervised analyses follow a class comparison paradigm, with the aim of discovering patterns of gene expression that support, or co-vary with assigned membership in new classes discovered in the unsupervised analysis, or with the clinical parameters.

25 Clinical data obtained through additional studies is the basis for potential classification groups, including pulmonary physiologic testing and demographic data. For each discriminator a binary class comparison analysis is conducted, and the reliability statistics (as determined by permutation testing) associated with the postulated classification are determined.

30 Once a signature or discriminator list of genes between the groups is determined, the expression data is evaluated using alternative techniques (Fig. 5). Quantitative reverse transcription-PCR (RT-PCR) is accepted as the "gold standard" validation method for array

data. If differentially expressed cell signaling transcripts are present, these findings are validated with flow cytometry cell marker experiments. This methodology not only validates the array findings, but provides useful information regarding the biology of any interesting candidate genes identified in the experiment. More extensive validation via protein abundance
5 measurements in PBMC or in other tissues are considered in light of the biological function of the biomarkers discovered.

The independent validation (Fig. 5) of the differentially expressed genes is directed by the signature expression profile for sample discrimination obtained as discussed above. Since this methodologic approach is determined by the observed result, this study is designed to
10 collect and maintain appropriate samples for future validation. Additional RNA is maintained from the original isolation for quantitative PCR of differentially expressed transcripts. Second, an aliquot of PBMC is cryopreserved at the time of initial phlebotomy for future analysis by either western blot or flow cytometry if a cell surface marker of interest is differentially expressed. Specific antibodies of interest pertaining to highly differentially expressed genes or
15 genes of biologic importance are used.

Reference List

1. Lopez, A. D. and C. C. Murray. 1998. The global burden of disease, 1990-2020. Nat.Med. 4:1241-1243.

20 2. U.S.Department of Health and Human Services, National Institutes of Health National Heart Lung and Blood Institute. NIH Publication No. 03-5229. 2003.

3. Barnes, P. J. 2000. Chronic obstructive pulmonary disease. N.Engl.J.Med. 343:269-280.

4. Fletcher, C. and R. Peto. 1977. The natural history of chronic airflow obstruction.
25 Br.Med.J. 1:1645-1648.

5. Rutgers, S. R., D. S. Postma, N. H. ten Hacken, H. F. Kauffman, T. W. Der Mark, G. H. Koeter, and W. Timens. 2000. Ongoing airway inflammation in patients with COPD who Do not currently smoke. Chest 117:262S.

6. Bullinger, L., K. Dohner, E. Bair, S. Frohling, R. F. Schlenk, R. Tibshirani, H. Dohner,
30 and J. R. Pollack. 2004. Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. N.Engl.J.Med. 350:1605-1616.

7. Valk, P. J., R. G. Verhaak, M. A. Beijen, C. A. Erapelink, Barjesteh van Waalwijk van Doorn-Khosrovani, J. M. Boer, H. B. Beverloo, M. J. Moorhouse, P. J. van der Spek, B. Lowenberg, and R. Delwel. 2004. Prognostically useful gene-expression profiles in acute myeloid leukemia. *N.Engl.J.Med.* 350:1617-1628.
- 5 8. Bovin, L. F., K. Rieneck, C. Workman, H. Nielsen, S. F. Sorensen, H. Skjodt, A. Florescu, S. Brunak, and K. Bendtzen. 2004. Blood cell gene expression profiling in rheumatoid arthritis. Discriminative genes and effect of rheumatoid factor. *Immunol.Lett.* 93:217-226.
9. Jison, M. L., P. J. Munson, J. J. Barb, A. F. Suffredini, S. Talwar, C. Logun, N. Raghavachari, J. H. Beigel, J. H. Shelhamer, R. L. Danner, and M. T. Gladwin. 2004. Blood
10 mononuclear cell gene expression profiles characterize the oxidant, hemolytic, and inflammatory stress of sickle cell disease. *Blood* 104:270-280.
10. Olsen, N. J., J. H. Moore, and T. M. Aune. 2004. Gene expression signatures for autoimmune disease in peripheral blood mononuclear cells. *Arthritis Res.Ther.* 6:120-128.
11. Ramanathan, M., B. Weinstock-Guttman, L. T. Nguyen, D. Badgett, C. Miller, K.
15 Patrick, C. Brownschidle, and L. Jacobs. 2001. In vivo gene expression revealed by cDNA arrays: the pattern in relapsing-remitting multiple sclerosis patients compared with normal subjects. *J.Neuroimmunol.* 116:213-219.
12. DePrimo, S. E., L. M. Wong, D. B. Khatry, S. L. Nicholas, W. C. Manning, B. D. Smolich, A. M. O'Farrell, and J. M. Cherrington. 2003. Expression profiling of blood samples
20 from an SU5416 Phase III metastatic colorectal cancer clinical trial: a novel strategy for biomarker identification. *BMC.Cancer* 3:3.
13. Bull, T. M., C. D. Coldren, M. Moore, S. M. Sotto-Santiago, D. V. Pham, S. P. Nana-Sinkam, N. F. Voelkel, and M. W. Geraci. 2004. Gene Microarray Analysis of Peripheral Blood Cells in Pulmonary Arterial Hypertension. *Am.J.Respir.Crit Care Med.*
- 25 14. Anthonisen, N. R., J. E. Connett, and R. P. Murray. 2002. Smoking and lung function of Lung Health Study participants after 11 years. *Am.J.Respir.Crit Care Med.* 166:675-679.
15. Sutherland, E. R. and R. M. Cherniack. 2004. Management of chronic obstructive pulmonary disease. *N.Engl.J.Med.* 350:2689-2697.

16. Silverman, E. K. and F. E. Speizer. 1996. Risk factors for the development of chronic obstructive pulmonary disease. *Med.Clin.North Am.* 80:501-522.
17. Redline, S., P. V. Tishler, F. I. Lewitter, I. B. Tager, A. Munoz, and F. E. Speizer. 1987. Assessment of genetic and nongenetic influences on pulmonary function. A twin study. *Am.Rev.Respir.Dis.* 135:217-222.
18. Tager, I. B., B. Rosner, P. V. Tishler, F. E. Speizer, and E. H. Kass. 1976. Household aggregation of pulmonary function and chronic bronchitis. *Am.Rev.Respir.Dis.* 114:485-492.
19. Barnes, P. J. 1999. Genetics and pulmonary medicine. 9. Molecular genetics of chronic obstructive pulmonary disease. *Thorax* 54:245-252.
20. Barnes, P. J. 2003. New concepts in chronic obstructive pulmonary disease. *Annu.Rev.Med.* 54:113-129.
21. Barnes, P. J. 2000. Mechanisms in COPD: differences from asthma. *Chest* 117:10S-14S.
22. Keatings, V. M., P. D. Collins, D. M. Scott, and P. J. Barnes. 1996. Differences in interleukin-8 and tumor necrosis factor-alpha in induced sputum from patients with chronic obstructive pulmonary disease or asthma. *Am.J.Respir.Crit Care Med.* 153:530-534.
23. Saetta, M., G. Turato, P. Maestrelli, C. E. Mapp, and L. M. Fabbri. 2001. Cellular and structural bases of chronic obstructive pulmonary disease. *Am.J.Respir.Crit Care Med.* 163:1304-1309.
24. Di Stefano, A., G. Turato, P. Maestrelli, C. E. Mapp, M. P. Ruggieri, A. Roggeri, P. Boschetto, L. M. Fabbri, and M. Saetta. 1996. Airflow limitation in chronic bronchitis is associated with T-lymphocyte and macrophage infiltration of the bronchial mucosa. *Am.J.Respir.Crit Care Med.* 153:629-632.
25. Jeffery, P. K. 1998. Structural and inflammatory changes in COPD: a comparison with asthma. *Thorax* 53:129-136.
26. Saetta, M., A. Di Stefano, P. Maestrelli, A. Ferrarresso, R. Drigo, A. Potena, A. Ciaccia, and L. M. Fabbri. 1993. Activated T-lymphocytes and macrophages in bronchial mucosa of subjects with chronic bronchitis. *Am.Rev.Respir.Dis.* 147:301-306.

27. Thompson, A. B., D. Daughton, R. A. Robbins, M. A. Ghafouri, M. Oehlerking, and S. I. Rennard. 1989. Intraluminal airway inflammation in chronic bronchitis. Characterization and correlation with clinical parameters. *Am.Rev.Respir.Dis.* 140:1527-1537.
28. O'Shaughnessy, T. C., T. W. Ansari, N. C. Barnes, and P. K. Jeffery. 1997.
5 Inflammation in bronchial biopsies of subjects with chronic bronchitis: inverse relationship of CD8+ T lymphocytes with FEV1. *Am.J.Respir.Crit Care Med.* 155:852-857.
29. Tzanakis, N., G. Chrysofakis, M. Tsoumakidou, D. Kyriakou, J. Tsiligianni, D. Bouros, and N. M. Siafakas. 2004. Induced sputum CD8+ T-lymphocyte subpopulations in chronic obstructive pulmonary disease. *Respir.Med.* 98:57-65.
- 10 30. Di Stefano, A., A. Capelli, M. Lusuardi, P. Balbo, C. Vecchio, P. Maestrelli, C. E. Mapp, L. M. Fabbri, C. F. Donner, and M. Saetta. 1998. Severity of airflow limitation is associated with severity of airway inflammation in smokers. *Am.J.Respir.Crit Care Med.* 158:1277-1285.
- 15 31. Majo, J., H. Ghezzi, and M. G. Cosio. 2001. Lymphocyte population and apoptosis in the lungs of smokers and their relation to emphysema. *Eur.Respir.J.* 17:946-953.
32. Saetta, M., S. Baraldo, L. Corbino, G. Turato, F. Braccioni, F. Rea, G. Cavallesco, G. Tropeano, C. E. Mapp, P. Maestrelli, A. Ciaccia, and L. M. Fabbri. 1999. CD8+ve cells in the lungs of smokers with chronic obstructive pulmonary disease. *Am.J.Respir.Crit Care Med.* 160:711-717.
- 20 33. Saetta, M., A. Di Stefano, G. Turato, F. M. Facchini, L. Corbino, C. E. Mapp, P. Maestrelli, A. Ciaccia, and L. M. Fabbri. 1998. CD8+ T-lymphocytes in peripheral airways of smokers with chronic obstructive pulmonary disease. *Am.J.Respir.Crit Care Med.* 157:822-826.
- 25 34. Liu, A. N., A. Z. Mohammed, W. R. Rice, D. T. Fiedeldej, J. S. Liebermann, J. A. Whitsett, T. J. Braciale, and R. I. Enelow. 1999. Perforin-independent CD8(+) T-cell-mediated cytotoxicity of alveolar epithelial cells is preferentially mediated by tumor necrosis factor-alpha: relative insensitivity to Fas ligand. *Am.J.Respir.Cell Mol.Biol.* 20:849-858.
35. Enelow, R. I., A. Z. Mohammed, M. H. Stoler, A. N. Liu, J. S. Young, Y. H. Lou, and T. J. Braciale. 1998. Structural and functional consequences of alveolar cell recognition by CD8(+) T lymphocytes in experimental lung disease. *J.Clin.Invest* 102:1653-1661.

36. Rahman, I., D. Morrison, K. Donaldson, and W. MacNee. 1996. Systemic oxidative stress in asthma, COPD, and smokers. *Am.J.Respir.Crit Care Med.* 154:1055-1060.
37. Burnett, D., A. Chamba, S. L. Hill, and R. A. Stockley. 1987. Neutrophils from subjects with chronic obstructive lung disease show enhanced chemotaxis and extracellular proteolysis. *Lancet* 2:1043-1046.
38. Noguera, A., S. Batle, C. Miralles, J. Iglesias, X. Busquets, W. MacNee, and A. G. Agusti. 2001. Enhanced neutrophil response in chronic obstructive pulmonary disease. *Thorax* 56:432-437.
39. Noguera, A., X. Busquets, J. Sauleda, J. M. Villaverde, W. MacNee, and A. G. Agusti. 1998. Expression of adhesion molecules and G proteins in circulating neutrophils in chronic obstructive pulmonary disease. *Am.J.Respir.Crit Care Med.* 158:1664-1668.
40. Noguera, A., E. Sala, A. R. Pons, J. Iglesias, W. MacNee, and A. G. Agusti. 2004. Expression of adhesion molecules during apoptosis of circulating neutrophils in COPD. *Chest* 125:1837-1842.
41. Di Francia, M., D. Barbier, J. L. Mege, and J. Orehek. 1994. Tumor necrosis factor-alpha levels and weight loss in chronic obstructive pulmonary disease. *Am.J.Respir.Crit Care Med.* 150:1453-1455.
42. Schols, A. M., W. A. Buurman, Staal van den Brekel AJ, M. A. Dentener, and E. F. Wouters. 1996. Evidence for a relation between metabolic derangements and increased levels of inflammatory mediators in a subgroup of patients with chronic obstructive pulmonary disease. *Thorax* 51:819-824.
43. de, G., I, M. Donahoe, W. J. Calhoun, J. Mancino, and R. M. Rogers. 1996. Elevated TNF-alpha production by peripheral blood monocytes of weight-losing COPD patients. *Am.J.Respir.Crit Care Med.* 153:633-637.
44. Miller, L. G., G. Goldstein, M. Murphy, and L. C. Ginns. 1982. Reversible alterations in immunoregulatory T cells in smoking. Analysis by monoclonal antibodies and flow cytometry. *Chest* 82:526-529.

45. Costabel, U., K. J. Bross, C. Reuter, K. H. Ruhle, and H. Matthys. 1986. Alterations in immunoregulatory T-cell subsets in cigarette smokers. A phenotypic analysis of bronchoalveolar and blood lymphocytes. *Chest* 90:39-44.
46. Ekberg-Jansson, A., B. Andersson, E. Avra, O. Nilsson, and C. G. Lofdahl. 2000. The expression of lymphocyte surface antigens in bronchial biopsies, bronchoalveolar lavage cells and blood cells in healthy smoking and never-smoking men, 60 years old. *Respir.Med.* 94:264-272.
47. de Jong, J. W., B. Belt-Gritter, G. H. Koeter, and D. S. Postma. 1997. Peripheral blood lymphocyte cell subsets in subjects with chronic obstructive pulmonary disease: association with smoking, IgE and lung function. *Respir.Med.* 91:67-76.
48. Kim, W. D., W. S. Kim, Y. Koh, S. D. Lee, C. M. Lim, D. S. Kim, and Y. J. Cho. 2002. Abnormal peripheral blood T-lymphocyte subsets in a subgroup of patients with COPD. *Chest* 122:437-444.
49. Hodge, S. J., G. L. Hodge, P. N. Reynolds, R. Scicchitano, and M. Holmes. 2003. Increased production of TGF-beta and apoptosis of T lymphocytes isolated from peripheral blood in COPD. *Am.J.Physiol Lung Cell Mol.Physiol* 285:L492-L499.
50. Takabatake, N., H. Nakamura, S. Inoue, K. Terashita, H. Yuki, S. Kato, S. Yasumura, and H. Tomoike. 2000. Circulating levels of soluble Fas ligand and soluble Fas in patients with chronic obstructive pulmonary disease. *Respir.Med.* 94:1215-1220.
51. Lehmann, C., A. Wilkening, D. Leiber, A. Markus, N. Krug, R. Pabst, and T. Tschernig. 2001. Lymphocytes in the bronchoalveolar space reenter the lung tissue by means of the alveolar epithelium, migrate to regional lymph nodes, and subsequently rejoin the systemic immune system. *Anat.Rec.* 264:229-236.
52. Hall, M. A., K. R. Ahmadi, P. Norman, H. Snieder, A. J. MacGregor, R. W. Vaughan, T. D. Spector, and J. S. Lanchbury. 2000. Genetic influence on peripheral blood T lymphocyte levels. *Genes Immun.* 1:423-427.
53. Amadori, A., R. Zamarchi, G. De Silvestro, G. Forza, G. Cavatton, G. A. Danieli, M. Clementi, and L. Chieco-Bianchi. 1995. Genetic control of the CD4/CD8 T-cell ratio in humans. *Nat.Med.* 1:1279-1283.

54. Whitney, A. R., M. Diehn, S. J. Popper, A. A. Alizadeh, J. C. Boldrick, D. A. Relman, and P. O. Brown. 2003. Individuality and variation in gene expression patterns in human blood. *Proc.Natl.Acad.Sci.U.S.A* 100:1896-1901.

55. Rus, V., S. P. Atamas, V. Shustova, I. G. Luzina, F. Selaru, L. S. Magder, and C. S. Via. 2002. Expression of cytokine- and chemokine-related genes in peripheral blood mononuclear cells from lupus patients by cDNA array. *Clin.Immunol.* 102:283-290.

56. Fabbri, L. M. and S. S. Hurd. 2003. Global Strategy for the Diagnosis, Management and Prevention of COPD: 2003 update. *Eur.Respir.J.* 22:1-2.

57. Hintze JL. PASS 2001. NCSS, Kaysville, Utah.

10

58. Yang, Y. H. and T. Speed. 2002. Design issues for cDNA microarray experiments. *Nat.Rev.Genet.* 3:579-588.

59. Dobbin K and Simon RM. 2005. Sample size determination in microarray experiments for class comparison and prognostic classification. *Biostatistics* 6: 27-38.

15 Each reference disclosed or cited herein is incorporated herein by reference in its entirety.

While various embodiments of the present invention have been described in detail, it is apparent that modifications and adaptations of those embodiments will occur to those skilled in the art. It is to be expressly understood, however, that such modifications and adaptations are within the scope of the present invention, as set forth in the following claims.

20

What is claimed is:

1. A method to diagnose chronic obstructive pulmonary disease (COPD) or a predisposition to develop COPD, comprising:

5 a) detecting the level of expression of at least one gene in a sample of peripheral blood cells isolated from a patient to be tested, wherein the gene is chosen from the genes represented by SEQ ID NO:1-323, and wherein the level of expression of each of the genes in any one or more of Tables 2-5 is associated with COPD as measured by either upregulation or downregulation of gene expression in peripheral blood cells from patients with COPD as compared to the level of expression of the genes in peripheral blood cells from normal controls; and

10 b) comparing the level of expression of the gene from the patient sample to the level of expression of the gene in normal control peripheral blood cells, wherein detection of regulation of the expression of the gene in the patient sample in the direction associated with COPD indicated in Table 2, 3, 4 and/or 5 indicates a diagnosis of COPD in the patient.

15 2. The method of Claim 1, wherein the step (a) of detecting comprises detecting expression of at least 5 genes chosen from the genes represented by SEQ ID NO:1-323.

3. The method of Claim 1, wherein the step (a) of detecting comprises detecting expression of at least 10 genes chosen from the genes represented by SEQ ID NO:1-323.

20 4. The method of Claim 1, wherein the step (a) of detecting comprises detecting expression of at least 15 genes chosen from the genes represented by SEQ ID NO:1-323.

5. The method of Claim 1, wherein the step (a) of detecting comprises detecting expression of at least 20 genes chosen from the genes represented by SEQ ID NO:1-323.

25 6. The method of Claim 1, wherein the step (a) of detecting comprises detecting expression of at least 25 genes chosen from the genes represented by SEQ ID NO:1-323.

7. The method of Claim 1, wherein the step (a) of detecting comprises detecting expression of at least 50 genes chosen from the genes represented by SEQ ID NO:1-323.

8. The method of Claim 1, wherein the step (a) of detecting comprises detecting expression of at least 75 genes chosen from the genes represented by SEQ ID NO:1-323.

30 9. The method of Claim 1, wherein the step (a) of detecting comprises detecting expression of at least 100 genes chosen from the genes represented by SEQ ID NO:1-229, and 320-324.

10. The method of Claim 1, wherein the step (a) of detecting comprises detecting expression of at least 125 genes chosen from the genes represented by SEQ ID NO:1-229, and 320-324.

11. The method of Claim 1, wherein the step (a) of detecting comprises detecting
5 expression of at least 150 genes chosen from the genes represented by SEQ ID NO:1-229, and 320-324.

12. The method of Claim 1, wherein the step (a) of detecting comprises detecting expression of at least 175 genes chosen from the genes represented by SEQ ID NO:1-229, and 320-324.

10 13. The method of Claim 1, wherein the step (a) of detecting comprises detecting expression of at least 200 genes chosen from the genes represented by SEQ ID NO:1-229, and 320-324.

14. The method of Claim 1, wherein the step (a) of detecting comprises detecting expression of at least 225 genes chosen from the genes represented by SEQ ID NO:1-229, and
15 320-324.

15. The method of Claim 1, wherein the step (a) of detecting comprises detecting expression of all of the genes represented by SEQ ID NO:1-323.

16. The method of any one of Claims 1-15, wherein expression of the gene is detected by measuring amounts of transcripts of the gene in the patient peripheral blood cells.

20 17. The method of any one of Claims 1-15, wherein expression of the gene is detected by detecting hybridization of at least a portion of the gene or a transcript thereof to a nucleic acid molecule comprising a portion of the gene or a transcript thereof in a nucleic acid array.

18. The method of any one of Claims 1-15, wherein expression of the gene is
25 detected by detecting the production of a protein encoded by the gene.

19. The method of any one of Claims 1-18, wherein the level of expression of the gene in the peripheral blood cells of a normal control has been predetermined.

20. A method to monitor the treatment of a patient with COPD, comprising:

30 a) detecting the level of expression of at least one gene in a sample of peripheral blood cells isolated from a patient undergoing treatment for COPD, wherein the gene is chosen from the genes represented by any one of SEQ ID NO:1-324, and wherein the level of expression of each of the genes represented by any one of SEQ ID NO:1-324 is associated with COPD as measured by either upregulation or downregulation of gene expression in peripheral blood cells from patients with COPD as

compared to the level of expression of the genes in peripheral blood cells from normal controls;

b) comparing the level of expression of the gene from the patient sample to the level of expression of the gene in a prior sample of peripheral blood cells from the patient, wherein detection of a change in the level of expression of the gene, as compared to the level of expression in the prior sample, toward the level of the expression of the gene in a normal control sample, indicates that the treatment for COPD is producing a beneficial result.

21. The method of Claim 20, wherein detection of a change in the level of expression of the gene, as compared to the level of expression in the prior sample, away from the level of the expression of the gene in a normal control sample, indicates a progression of the COPD.

22. The method of Claim 20, wherein detection of no significant change in the level of expression of the gene, as compared to the level of expression in the prior sample, indicates no significant change in the progression or treatment of the COPD in the patient.

23. A plurality of polynucleotides for the detection of the expression of genes that are indicative of COPD in a patient or a preclinical disposition therefore;

wherein the plurality of polynucleotides consists of at least two polynucleotides, wherein each polynucleotide is at least 5 nucleotides in length, and wherein each polynucleotide is complementary to an RNA transcript, or nucleotide derived therefrom, of a gene that is regulated differently in individuals with COPD as compared to individuals that do not have COPD.

24. The plurality of polynucleotides of Claim 23, wherein each polynucleotide is complementary to an RNA transcript, or a polynucleotide derived therefrom, of a gene represented by any one of SEQ ID NO:1-324.

25. The plurality of polynucleotides of Claim 23, wherein the plurality of polynucleotides comprises polynucleotides that are complementary to an RNA transcript, or a nucleotide derived therefrom, of at least two genes represented by any one of SEQ ID NO:1-324.

26. The plurality of polynucleotides of Claim 23, wherein the plurality of polynucleotides comprises polynucleotides that are complementary to an RNA transcript, or a nucleotide derived therefrom, of at least five genes represented by any one of SEQ ID NO:1-324.

27. The plurality of polynucleotides of Claim 23, wherein the plurality of polynucleotides comprises polynucleotides that are complementary to an RNA transcript, or a nucleotide derived therefrom, of at least 10 genes represented by any one of SEQ ID NO:1-324.

28. The plurality of polynucleotides of Claim 23, wherein the plurality of polynucleotides comprises polynucleotides that are complementary to an RNA transcript, or a nucleotide derived therefrom, of at least 50 genes represented by any one of SEQ ID NO:1-324.

29. The plurality of polynucleotides of Claim 23, wherein the plurality of polynucleotides comprises polynucleotides that are complementary to an RNA transcript, or a nucleotide derived therefrom, of at least 100 genes represented by any one of SEQ ID NO:1-229, and 320-324.

30. The plurality of polynucleotides of Claim 23, wherein the plurality of polynucleotides comprises polynucleotides that are complementary to an RNA transcript, or a nucleotide derived therefrom, of at least 150 genes represented by any one of SEQ ID NO:1-229, and 320-324.

31. The plurality of polynucleotides of Claim 23, wherein the plurality of polynucleotides comprises polynucleotides that are complementary to an RNA transcript, or a nucleotide derived therefrom, of at least 200 genes represented by any one of SEQ ID NO:1-229, and 320-324.

32. The plurality of polynucleotides of any one of Claims 23-31, wherein said polynucleotides are immobilized on a substrate.

33. The plurality of polynucleotides of any one of Claims 23-31, wherein said polynucleotides are hybridizable array elements in a microarray.

34. The plurality of polynucleotides of any one of Claims 23-31, wherein said polynucleotides are conjugated to detectable markers.

35. A method to diagnose chronic obstructive pulmonary disease (COPD) in a patient, comprising:

a) detecting the level of expression of at least one gene in a sample of peripheral blood cells isolated from a patient to be tested, wherein the gene is chosen from a group of genes, each of which has been previously identified to be upregulated or downregulated in the peripheral blood cells of patients who have been diagnosed with COPD, as compared to the level of expression of the gene in normal control peripheral blood cells; and

b) comparing the level of expression of the gene from the patient sample to the level of expression of the gene in normal control peripheral blood cells, wherein

detection of regulation of the expression of the gene in the patient sample in the direction associated with COPD as indicated by the previous identification, indicates a diagnosis of COPD in the patient.

5 36. A method to identify a compound with the potential to treat or prevent chronic obstructive pulmonary disease (COPD), comprising:

a) contacting a test compound with a cell that expresses a gene selected from any one or more of the genes represented by any one of SEQ ID NO:1-324;

10 b) identifying compounds that increase the expression or activity of genes represented by any one of SEQ ID NO:1-324 or the proteins encoded thereby that are downregulated in peripheral blood cells of patients with COPD as compared to peripheral blood cells of normal controls, or that decrease the expression or activity of genes represented by any one of SEQ ID NO:1-324 or the proteins encoded thereby that are upregulated in peripheral blood cells of patients with COPD as compared to peripheral blood cells of normal controls.

15 37. A method to treat a patient with COPD, comprising administering to the patient a therapeutic composition comprising a compound identified by the method of Claim 36.

38. The method of any or Claim 1, Claim 20, or 35, wherein detection of a change in the level of expression of at least one gene comprises detecting the presence of a protein.

20 39. The method of Claim 38, wherein the method further comprises detecting the presence of the protein using a reagent that specifically binds to the protein.

40. The method of claim 39, wherein the reagent is selected from the group consisting of an antibody, an antibody derivative, and an antibody fragment.

41. A plurality of reagents for the detection of the expression of genes that are indicative of COPD in a patient or a preclinical disposition therefore;

25 wherein the plurality of reagents consists of at least two reagents that each of which specifically bind to a protein, wherein each protein is at least 15 amino acids in length, and wherein each protein is encoded a gene that is regulated differently in individuals with COPD as compared to individuals that do not have COPD.

30 42. The plurality of reagents of Claim 41, wherein each protein is encoded a gene represented by any one of SEQ ID NO:1-324.

43. The plurality of reagents of Claim 41, wherein at least two proteins are encoded a gene represented by any one of SEQ ID NO:1-324.

44. The plurality of reagents of Claim 41, wherein at least five proteins are encoded a gene represented by any one of SEQ ID NO:1-324.

45. The plurality of reagents of Claim 41, wherein at least 10 proteins are encoded a gene represented by any one of SEQ ID NO:1-324..

46. The plurality of reagents of Claim 41, wherein at least 50 proteins are encoded a gene represented by any one of SEQ ID NO:1-324.

5 47. The plurality of reagents of Claim 41, wherein at least 100 proteins are encoded a gene represented by any one of SEQ ID NO:1-229, and 320-324.

48. The plurality of reagents of Claim 41, wherein at least 150 proteins are encoded a gene represented by any one of SEQ ID NO:1-229, and 320-324.

10 49. The plurality of reagents of Claim 41, wherein at least 200 proteins are encoded a gene represented by any one of SEQ ID NO:1-229, and 320-324.

50. The plurality of reagents of any one of Claims 41-49, wherein said reagents are immobilized on a substrate.

15 51. The plurality of reagents of any one of Claims 41-49, wherein each of said reagents are selected from the group consisting of an antibody, an antibody derivative, and an antibody fragment, and wherein each of said reagents are elements in a microarray.

52. The plurality of reagents of any one of Claims 41-49, wherein said reagents are conjugated to detectable markers.

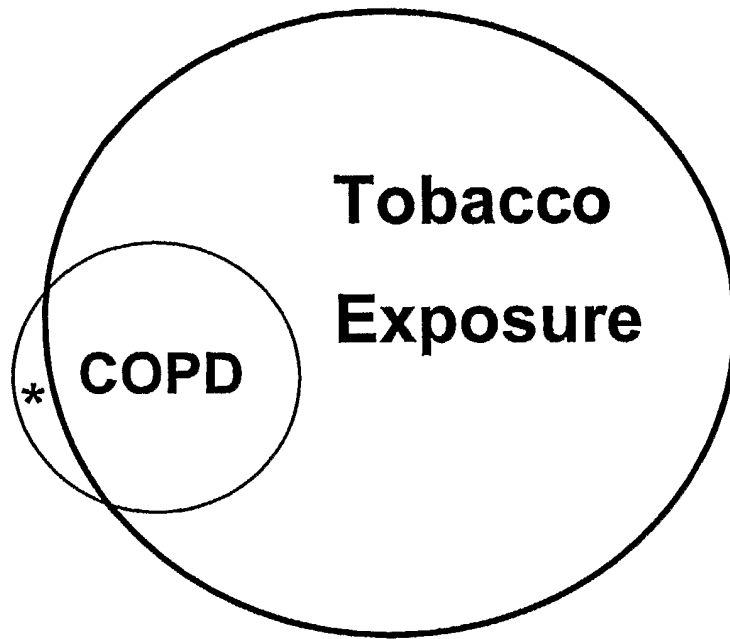


Figure 1

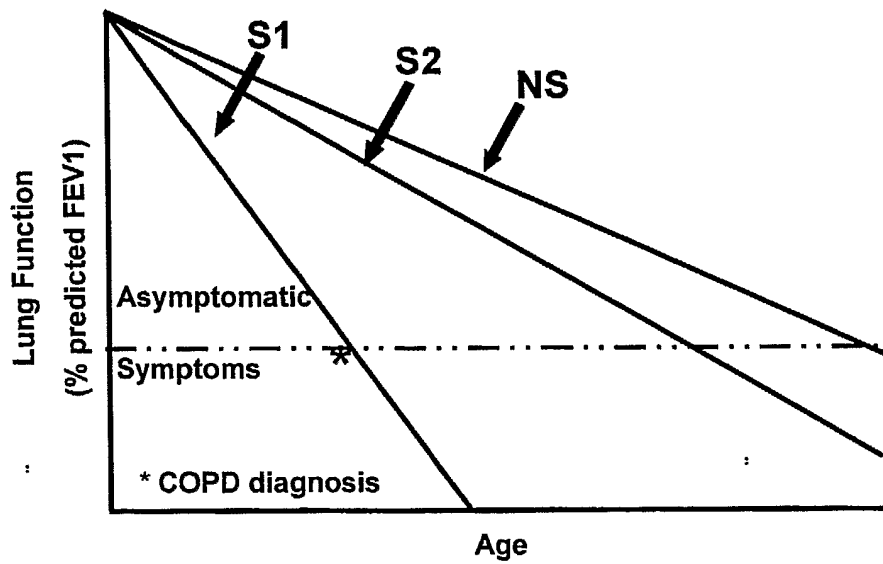


Figure 2

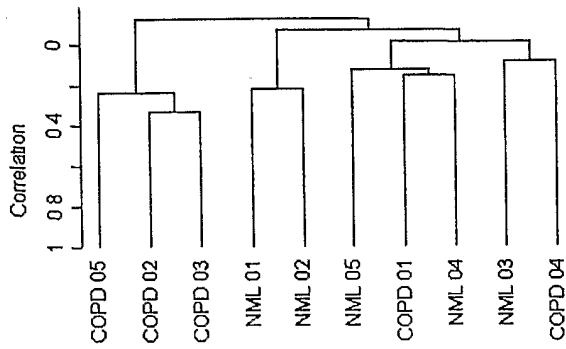


Figure 3A

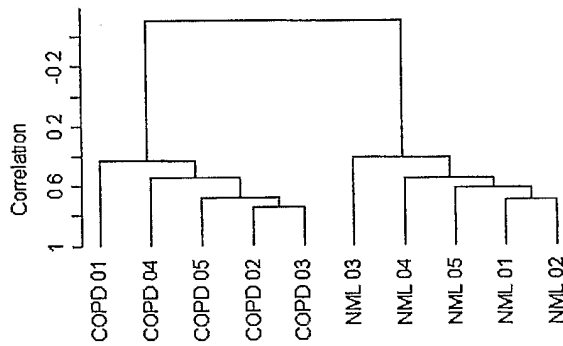


Figure 3B

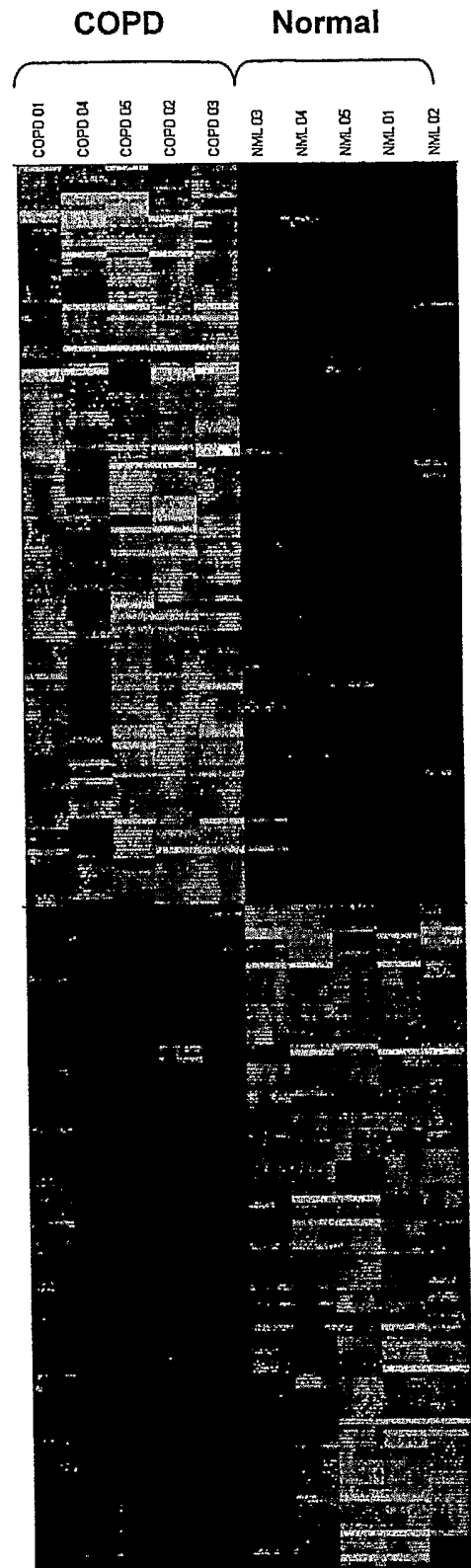


Figure 3C

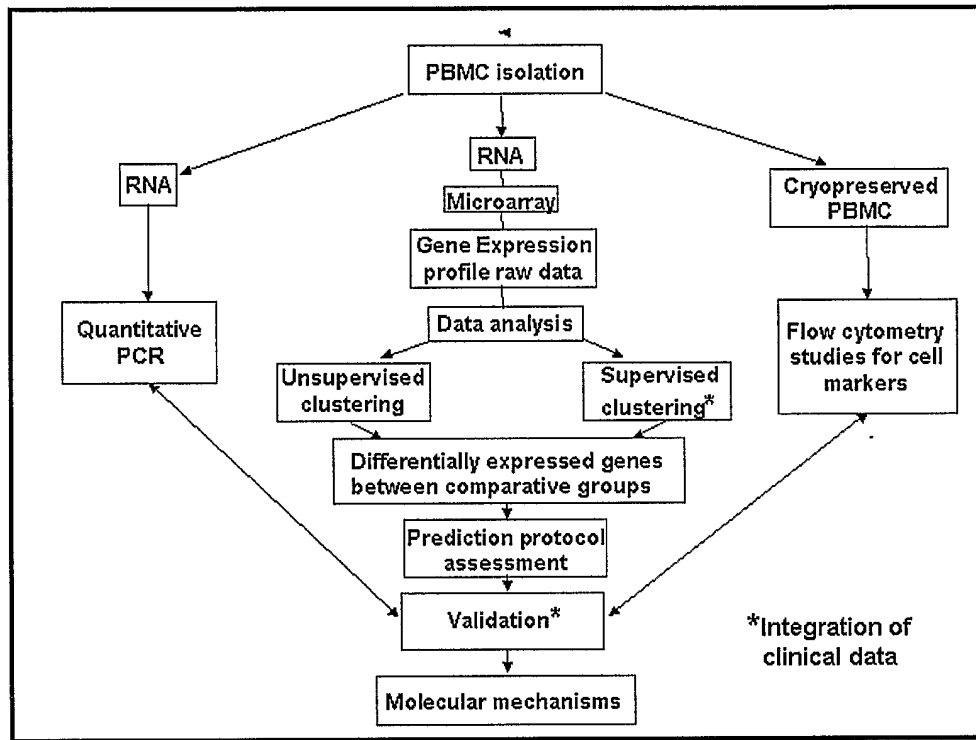


Figure 4

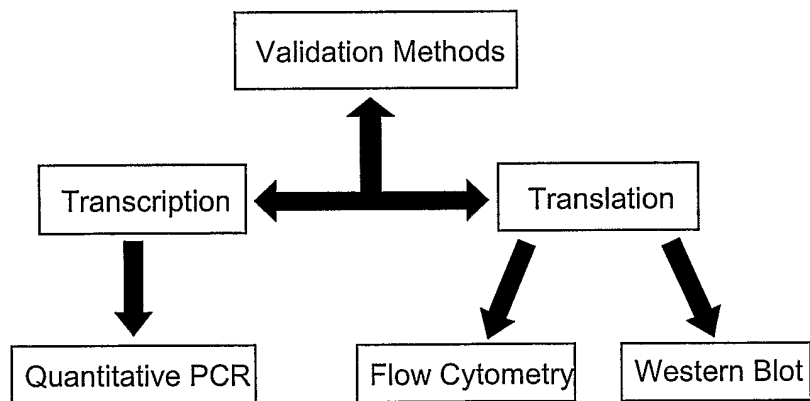


Figure 5