



(51) International Patent Classification:

G01N 33/68 (2006.01)

(21) International Application Number:

PCT/EP2023/062030

(22) International Filing Date:

05 May 2023 (05.05.2023)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

22305676.3

06 May 2022 (06.05.2022)

EP

(71) Applicants: **INSTITUT NATIONAL DE LA SANTÉ ET DE LA RECHERCHE MÉDICALE** [FR/FR]; 101, rue de Tolbiac, 75013 Paris (FR). **MEDIMMUNE, LLC** [US/US]; 1 Medimmune Way, Gaithersburg, Maryland 20878 (US). **UNIVERSITÉ PARIS CITÉ** [FR/FR]; 85 Boulevard Saint-Germain, 75006 Paris (FR).

(72) Inventors: **DAGHER, Rania**; MedImmune, LLC, 1 MedImmune Way, Gaithersburg, Maryland 20878 (US). **KOLBECK, Roland**; MedImmune, LLC 1 MedImmune Way, Gaithersburg, Maryland 20878 (US). **HUMBLES, Alison H.**; MedImmune, LLC 1 MedImmune Way, Gaithersburg, Maryland 20878 (US). **AUBIER, Michel**; FACULTE DE MEDECINE SITE XAVIER BICHAT - U1152 16 RUE HENRI HUCHARD, 75018 PARIS (FR).

(74) Agent: **INSERM TRANSFERT**; 7 rue Watt, 75013 Paris (FR).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO,

RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

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

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))
- with sequence listing part of description (Rule 5.2(a))
- with information concerning incorporation by reference of missing element or part (Rule 20.6)
- in black and white; the international application as filed contained color or greyscale and is available for download from PATENTSCOPE

(54) Title: METHODS FOR PREDICTING THE CLINICAL OUTCOME OF PATIENTS SUFFERING FROM CHRONIC OBSTRUCTIVE PULMONARY DISEASE

(57) Abstract: Chronic obstructive pulmonary disease (COPD) is a heterogeneous disease with pulmonary and extra-pulmonary manifestations. COPD heterogeneity is poorly understood and insights into molecular mechanisms is required to provide better treatment with targeted therapies. The inventors collected serum samples from (241) COPD subjects in the COBRA cohort and measured the expression of (1305) proteins using SOMAscan proteomic platform. Modular analyses and clustering of the proteomics were applied to identify disease subtypes. Cluster discoveries were revalidated during a follow up visit, and confirmed in a different COPD cohort. Unsupervised clustering using protein modules identified two clusters within COPD subjects. One cluster presented a higher expression of proteins associated with enhanced immune and pro-survival activities, host defense and cell renewal. Subjects in this cluster had a lower incidence of exacerbations, unscheduled medical visits, emphysema and diabetes. These protein signatures were conserved during a follow up visit, and validated in another COPD cohort. Among the 96 proteins, the inventors identified a small specific signature consisting of 15 proteins that were able to accurately differentiate the two patient clusters. This signature would thus be useful for predicting the clinical outcome of patients suffering from COPD.

METHODS FOR PREDICTING THE CLINICAL OUTCOME OF PATIENTS SUFFERING FROM CHRONIC OBSTRUCTIVE PULMONARY DISEASE

5 FIELD OF THE INVENTION:

The present invention is in the field of medicine, in particular pneumology.

BACKGROUND OF THE INVENTION:

Chronic obstructive pulmonary disease (COPD) is a difficult to treat disease, characterised by
10 irreversible airflow obstruction and often associated with lung emphysema. These events result
from persistent airway and lung inflammation and tissue remodelling that leads to respiratory
insufficiency and functional disability (1). Cigarette smoke, but also genetic/epigenetic
alterations leading to lung accelerated aging, have been shown to predispose individuals to
COPD (2–4). However, whether markers of these processes are detected in peripheral blood of
15 patients and relate to clinical traits of the disease remains elusive.

COPD manifests in different clinical phenotypes according to the degree of airflow obstruction,
frequency of acute exacerbations, emphysema and airway inflammation (5). In addition,
pulmonary and cardio-metabolic comorbidities may impact COPD prognosis and therapeutic
20 management (6). Therefore, numerous studies have addressed analytical approaches for
identifying novel COPD endotypes underlying these clinical phenotypes (7–9). Hence, specific
markers were established for identifying patients with high versus low rate of exacerbations
(10), or by defining COPD-subtypes using quantitative computed tomography (CT) imaging,
namely, emphysema- and airway-dominant diseases (11,12). However, these generated
25 profiles, have not been supported across multiple cohorts. Other approaches, which still
requiring validation, include the assessment of potential links between clinical traits of COPD
and changes in lung microbiome, systemic inflammation and pulmonary, or extra-pulmonary
co-morbidities (7,9).

30 Given the heterogeneity in COPD pathophysiology and clinical presentation, robust and wide
analytic approaches are required to characterize novel endotypes. To this end, High-throughput
proteomic technology, Slow Off-rate Modified Aptamers (SOMA)scan, has been developed for
quantitatively assessing hundreds of proteins in serum, plasma and tissue samples with high
sensitivity and specificity (13,14). This platform has previously proved useful in being able to

measure simultaneously large numbers of proteins in studies investigating novel biomarkers for organ diseases, including lung (14–16).

5 SUMMARY OF THE INVENTION:

The present invention is defined by the claims. In particular, the present invention relates to methods for predicting the clinical outcome of patients suffering from chronic obstructive pulmonary disease (COPD).

10 DETAILED DESCRIPTION OF THE INVENTION:

Chronic obstructive pulmonary disease (COPD) is a heterogeneous disease with pulmonary and extra-pulmonary manifestations. COPD heterogeneity is poorly understood and insights into molecular mechanisms is required to provide better treatment with targeted therapies.

The objectives of the inventors were to systematically profile serum proteins in COPD subjects, and to study disease mechanisms and molecular heterogeneity in relation to clinical presentation.

The inventors collected serum samples from 241 COPD subjects in the COBRA cohort and measured the expression of 1305 proteins using SOMAscan proteomic platform. Modular analyses and clustering of the proteomics were applied to identify disease subtypes. Functional annotation of the subtypes and association with key clinical parameters were examined. Cluster discoveries were revalidated during a follow up visit, and confirmed in a different COPD cohort.

Two clusters were identified within COPD subjects at inclusion. Cluster 1 showed elevated levels of some factors contributing to tissue injury, whereas Cluster 2 had higher expression of proteins associated with enhanced immune and defense activities, cell fate, remodelling and repair and altered metabolism and mitochondrial functions. Patients in Cluster 2 had a lower incidence of exacerbations, unscheduled medical visits and prevalence of emphysema and diabetes. These protein expression patterns were conserved, by a large part, during a second visit, and validated in another COPD patient group. A signature of 15 proteins that differentiated the two COPD clusters at the 2 visits and that enabled tracking of patients who switched cluster at their second visit was also identified. Overexpression of factors involved in host defense, pro-survival and cell renewal contribute to better clinical outcomes in COPD. In addition, high throughput proteomic assay proved usefulness for investigating molecular mechanisms

involved in COPD and for identifying targeted therapies. This signature would thus be useful for predicting the clinical outcome of patients suffering from COPD.

Main definitions:

5

As used herein, the term “**chronic obstructive pulmonary disease**” or “**COPD**” has its general meaning in the art and refers to and refers to a set of physiological symptoms including chronic cough, expectoration, exertional dyspnea and a significant, progressive reduction in airflow that may or may not be partly reversible. COPD is a disease characterized by a progressive airflow limitation caused by an abnormal inflammatory reaction to the chronic inhalation of particles. The Global Initiative for Chronic Obstructive Lung Disease (GOLD) has classified 4 different stages of COPD (**Table A**). In some embodiments, the patient suffers from moderate COPD. In some embodiments, the patient suffers from a severe or very severe COPD.

10

15 **Table A: Gold classification:** The volume in a one-second forced exhalation is called the forced expiratory volume in one second (FEV₁), measured in liters. The total exhaled breath is called the forced vital capacity (FVC), also measured in liters. In people with normal lung function, FEV₁ is at least 70% of FVC.

Stage I	Mild COPD	FEV ₁ /FVC<0.70	FEV ₁ ≥ 80% normal
Stage II	Moderate COPD	FEV ₁ /FVC<0.70	FEV ₁ 50-79% normal
Stage III	Severe COPD	FEV ₁ /FVC<0.70	FEV ₁ 30-49% normal
Stage IV	Very Severe COPD	FEV ₁ /FVC<0.70	FEV ₁ <30% normal, or <50% normal with chronic respiratory failure present

20 As used herein, the term “**clinical outcome**” has its general meaning in the art and refers to the health status of a patient following treatment for a disease or disorder, or in the absence of treatment. Clinical outcomes include, but are not limited to, an increase in the length of time until death, a decrease in the length of time until death, an increase in the chance of survival, an increase in the risk of death, survival, disease-free survival, advanced disease, death, and
25 favorable, poor response to therapy and worsening.

As used herein, the term “**worsening**” means that the COPD evolves at a later stage with respect to the first measured time point phase. The person skilled in the art will recognize and confirm

if the evolution of the disease is worse by also analyzing other indicative characteristics. The characteristics or indications that appear at a later stage, and that are indicative of a disease progression are, without limitation, the incidence of acute exacerbations, the occurrence of unscheduled medical visits, and the occurrence of hospitalizations for COPD within the previous year and emphysema. In particular, the worsening may also include the occurrence of co-morbidities, particularly diabetes and obstructive sleep apnea, as well as the need of medications for COPD (long-acting beta2-agonists, long-acting anti-muscarinic antagonists, inhaled corticosteroids...).

As used herein, the term “**risk of worsening**” is to be understood as referring to the probability of worsening.

As used herein, the term “**risk**” in the context of the present invention, relates to the probability that an event will occur over a specific time period and can mean a subject's “absolute risk” or “**relative risk**”. Absolute risk can be measured with reference to either actual observation post-measurement for the relevant time cohort, or with reference to index values developed from statistically valid historical cohorts that have been followed for the relevant time period. Relative risk refers to the ratio of absolute risks of a subject compared either to the absolute risks of low risk cohorts or an average population risk, which can vary by how clinical risk factors are assessed. Odds ratios, the proportion of positive events to negative events for a given test result, are also commonly used (odds are according to the formula $p/(1-p)$ where p is the probability of event and $(1-p)$ is the probability of no event) to no- conversion. “**Risk evaluation**,” or “**evaluation of risk**” in the context of the present invention encompasses making a prediction of the probability, odds, or likelihood that an event or disease state may occur, the rate of occurrence of the event or conversion from one disease state to another. Risk evaluation can also comprise prediction of future clinical parameters, traditional laboratory risk factor values, or other indices of relapse, either in absolute or relative terms in reference to a previously measured population. The methods of the present invention may be used to make continuous or categorical measurements of the risk of conversion, thus diagnosing and defining the risk spectrum of a category of subjects defined as being at risk of conversion. In the categorical scenario, the invention can be used to discriminate between normal and other subject cohorts at higher risk.

As used herein the term “**blood sample**” means any blood sample derived from the subject. Collections of blood samples can be performed by methods well known to those skilled in the art. In some embodiments, the blood sample is a serum sample or a plasma sample.

- 5 As used herein, the term “**BTK**” has its general meaning in the art and refers to the tyrosine-protein kinase BTK. The term is also known as agammaglobulinemia tyrosine kinase, B-cell progenitor kinase, and Bruton tyrosine kinase. An exemplary amino acid sequence for BTK is shown as SEQ ID NO:1.

10 SEQ ID NO:1 >sp|Q06187|BTK_HUMAN Tyrosine-protein kinase BTK OS=Homo sapiens OX=9606 GN=BTK PE=1 SV=3
 MAAVILESIFLKRSQQKKKTSPLNFKKRLFLLTVHKLSYYEYDFERGRSGSKGSIDVEK
 ITCVETVVPEKNPPPERQIPRRGEESSEMEQISIIERFPYPFQVVYDEGPLYVFSPTTEEL
 15 RKRWIHQKLVIRYNSDLVQKYHPCFWIDGQYLCCSQTAKNAMGCQILENRNGSLKPGSS
 HRKTKKPLPPTPEEDQILKKPLPPEPAAAPVSTSELKKVVALYDYMPMNANDLQLRKGDE
 YFILEESNLPWWRRARDKNGQEGYIPSNYVTEAEDSIEMYEWYSKHMTRSQAQQLLKQEGK
 EGGFIVRDSSKAGKYTVSVFAKSTGDPQGVIRHYVVCSTPQSQYYLAEKHLFSTIPELIN
 YHQHNSAGLISRLKYPVSQQNKNAPSTAGLGYGSWEIDPKDLTFLKELGTGQFGVVKYGK
 20 WRGQYDVAIKMIKEGSMSEDEFIEEAKVMMNLSHEKLVQLYGVCTKQRPFIITEYMANG
 CLLNYLREMRHRFQTQQLLEMCKDVCEAMEYLESKQFLHRDLAARNCLVNDQGVVKVSD
 GLSRYVLDDEYTSSVSGSKFPVRWSPPEVLMYSKFSSKSDIWAFAFVLMWEIYSLGKMPYER
 FTNSETAEHIAQGLRLRPLHLASEKVYTIMYSCWHEKADERPTFKILLSNILDVMDEES

- As used herein, the term “**Cyclophilin F**” has its general meaning in the art and refers to the
 25 peptidyl-prolyl cis-trans isomerase F, mitochondrial. The term is also known as PPIF or mitochondrial cyclophilin. An exemplary amino acid sequence for Cyclophilin F is shown as SEQ ID NO:2.

30 SEQ ID NO:2 >sp|P30405|PPIF_HUMAN Peptidyl-prolyl cis-trans isomerase F, mitochondrial OS=Homo sapiens OX=9606 GN=PPIF PE=1 SV=1
 MLALRCGSRWLGLLSVPRSVPLRLPAARACSKGSGDPSSSSSSGNPLVYLDVDANGKPLG
 RVVLELKADVVPKTAENFRALCTGEKGFGYKGSTFHRVIPSEFMCQAGDFTNHNGTGKSKI
 YGSRFPDENFTLKHVGPVLSMANAGPNTNGSQFFICTIKTDWLDGKHVVFHGHVKEGMDV
 35 VKKIESFGSKSGRTSKKIVITDCGQLS

As used herein, the term “**Tropomyosin 4**” has its general meaning in the art and refers to the tropomyosin alpha-4 chain. The term is also known as TPM4 or TM30p1. An exemplary amino acid sequence for Tropomyosin 4 is shown as SEQ ID NO:3.

40 SEQ ID NO:3 >sp|P67936|TPM4_HUMAN Tropomyosin alpha-4 chain OS=Homo sapiens OX=9606 GN=TPM4 PE=1 SV=3
 MAGLNSLEAVKRKIQALQQQADEADRAQGLQRELDGERERREKAEGDVAALNRRIQLVE
 EELDRAQERLATALQKLEEAKEADESERGMKVIENRAMKDEEKMEIQEMQLKEAKHIAE
 EADRYEEVARKLVILEGELERAEEAEVSELKCGDLEELKNVTNNLKSLEAASEKYSE
 45 KEDKYEEEIKLLSDKLKEAETRAEFAERTVAKLEKTIDDLKLAQAKEENVGLHQTLNQ

TLNELNCI

As used herein, the term “**Carbonic anhydrase XIII**” has its general meaning in the art and refers to the carbonic anhydrase 13. An exemplary amino acid sequence for carbonic anhydrase
 5 XIII is shown as SEQ ID NO:4.

SEQ ID NO:4>sp|Q8N1Q1|CAH13_HUMAN Carbonic anhydrase 13 OS=Homo sapiens
 OX=9606 GN=CA13 PE=1 SV=1
 MSRLSWGYPREHNGPIHWKEFFPIADGDQQSPIEIKTKEVKYDSSLRPLSIKYDPSSAKII
 10 SNSGHSFNVDFFDTENKSVLRGGPLTGSYRLRQVHLHWGSADDHGSEHIVDGVSYAAELH
 VVHWNSDKYPSFVEAAHEPDGLAVLGVFLQIGEPNSQLQKITDTLDSIKEKGKQTRFTNF
 DLLSLLPPSWDYWTYPGSLTVPPLLESVTWIVLKQPINISSQQLAKFRSLLCTAEGEAAA
 FLVSNHRPPQPLKGRKVRASFH

As used herein, the term “**FGF-16**” has its general meaning in the art and refers to the fibroblast
 15 growth factor 16. An exemplary amino acid sequence for FGF-16 is shown as SEQ ID NO:5.

SEQ ID NO:5>sp|O43320|FGF16_HUMAN Fibroblast growth factor 16 OS=Homo
 sapiens OX=9606 GN=FGF16 PE=1 SV=1
 20 MAEVGGVFASLDWDLHGFSSSLGNVPLADSPGFLLNERLGQIEGKLQRGSPDTFAHLKGIL
 RRRQLYCRTGFHLEIFPNGTVHGTNRDHSRFGILEFISLAVGLISIRGVDSGLYLGMMNER
 GELYGSKKLTRECVFREQFEENWYNTYASTLYKHSDSERQYYVALNKDGSREGYRTRKH
 QKFTHFLPRPVDPSKLPMSRDLFHYR

As used herein, the term “**AREG**” has its general meaning in the art and refers to the
 25 Amphiregulin. An exemplary amino acid sequence for AREG is shown as SEQ ID NO:6.

SEQ ID NO:6>sp|P15514|AREG_HUMAN Amphiregulin OS=Homo sapiens OX=9606
 30 GN=AREG PE=1 SV=2
 MRAPLLPPAPVVLSSLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG
 SEISPVSEMPSSSEPPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT
 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFICIGECKYIEHLEAVTCKCQQEYFGER
 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK
 35 LRQENGNVHAIA

As used herein, the term “**SHC1**” has its general meaning in the art and refers to the SHC-
 transforming protein 1. An exemplary amino acid sequence for SHC1 is shown as SEQ ID
 NO:7.

SEQ ID NO:7>sp|P29353|SHC1_HUMAN SHC-transforming protein 1 OS=Homo sapiens
 OX=9606 GN=SHC1 PE=1 SV=4
 40 MDLLPPKPKYNPLRNESLSSLEEGASGSTPPEELSPSPASSLGPIPLPPLPGDDSPPTLCS
 FFPRMSNLRLANPAGGRPGSKGEPGRAADDGEGIVGAAMPDSGPLPLQLQDMNKLSGGGGR
 45 RTRVEGGQLGGEWTRHGSFVNKPTRGWLHPNDKVMGPGVSYLVRYMGCVEVLQSMRALD
 FNTRTQVTREAIISLVCEAVPGAAGATRRRKPCSRPLSSILGRSNLKFAGMPITLTVSTSS

LNLMADCKQIIANHHMQSISFASGGDPDTAEYVAYVAKDPVNQRACHILECPEGLAQDV
 ISTIGQAFELRFKQYLRNPPKLVTPHDRMAGFDGSAWDEEEEEPPDHQYYNDFPGKEPPL
 GGVDMLRLREGAAPGAARPTAPNAQTPSHLGATLPVGQPVGGDPEVRKQMPPPPPCPGRE
 LFDDPSYVNVQNLDKARQAVGGAGPPNPAINGSAPRDLFDMKPFEDALRVPPPPQSVSMA
 5 EQLRGEPWFHKGKLSRREAEALLQLNGDFLVRESTTTTPGQYVLTGLQSGQPKHLLLVDEPG
 VVRTKDHREFESVSHLISYHMDNHLPIISAGSELCLQQPVERKL

As used herein, the term “**14-3-3 protein β/α** ” has its general meaning in the art and refers to the 14-3-3 protein beta/alpha. An exemplary amino acid sequence for 14-3-3 protein β/α is shown as SEQ ID NO:8.

SEQ ID NO:8>sp|P31946|1433B_HUMAN 14-3-3 protein beta/alpha OS=Homo sapiens
 OX=9606 GN=YWHAB PE=1 SV=3
 MTMDKSELVQAKLAEQAERYDDMAAMKAVTEQGHELSNEERNLLSVAYKNVVGARRSS
 15 WRVISSIEQKTERNEKKQQMKEYREKIEAELQDICNDVLELLDKYLIPNATQPESKVFY
 LKMKGDYFRYLSEVASGDNKQTTVSNSQQAYQEAFFEISKEMQPTHPIRLGLALNFSVFY
 YEILNSPEKACSLAKTAFDEAIAELDTLNEESYKDSTLIMQLLRDNLTLWTSENQGDEGD
 AGEGEN

As used herein, the term “**eIF-4H**” has its general meaning in the art and refers to the eukaryotic translation initiation factor 4H. An exemplary amino acid sequence for eIF-4H is shown as SEQ ID NO:9.

SEQ ID NO:9>sp|Q15056|IF4H_HUMAN Eukaryotic translation initiation factor
 4H OS=Homo sapiens OX=9606 GN=EIF4H PE=1 SV=5
 25 MADFDTYDDRAYSSFSGGRGSRGSAGGHGSRSQKELPTEPPYTAYVGNLPFNTVQGDIDA
 IFKDLISIRSVRLVRDKDIDKFKGFCYVEFDEVDSLKEALTYDGALLGDRSLRVDAIEGRK
 QDKGGFGFRKGPPDDRGMGSSRESRGWDSRDDFNSGFRDDFLGGRGGSRPGRRTGPPM
 GSRFRDGPPLRGSNMDFREPTTEERAQRRLQLKPRTVATPLNQVANPNSAIFGGARPRE
 30 EVVQKEQE

As used herein, the term “**CD40 ligand**” has its general meaning in the art and is also known as CD154 or CD40L. An exemplary amino acid sequence for CD40 ligand is shown as SEQ ID NO:10.

SEQ ID NO:10>sp|P29965|CD40L_HUMAN CD40 ligand OS=Homo sapiens OX=9606
 GN=CD40LG PE=1 SV=1
 MIETYNQTSRPAATGLPISMKIFMYLLTVFLITQMIGSALFAVYLHRRLDKIEDERNLH
 EDFVFMKTIQRCNTGERSLSLLNCEEIKSQFEGFVKDIMLNKEETKKENSFEMQKGDQNP
 40 QIAAHVISEASSKTTSVLQWAEKGYYTMSNNLVLTLENGKQLTVKRQGLYYIYAQVTFCSN
 REASSQAPFIASLCLKSPGRFERILLRAANTHSSAKPCGQQSIHLGGVFELQPGASVFN
 VTDPSQVSHGTGFTSFGLLKL

As used herein, the term “**H2A3**” has its general meaning in the art and refers to the stone H2A type 3. An exemplary amino acid sequence for H2A3 is shown as SEQ ID NO:11.

SEQ ID NO:11>sp|Q7L7L0|H2A3_HUMAN Histone H2A type 3 OS=Homo sapiens
 OX=9606 GN=H2AW PE=1 SV=3
 MSGRGKQGGKARAKAKSRSSRAGLQFPVGRVHRLLRKGNYSERVGAGAPVYLAADVLEYLT
 AEILELAGNAARDNKKTRIIPRHLQLAIRNDEELNKKLLGRVTIAQGGVLPNIQAVLLPKK
 TESHKAKGK

As used herein, the term “**Midkine**” has its general meaning in the art and is also known as
 NRG2 (« neurite growth-promoting factor 2 »). An exemplary amino acid sequence for
 Midkine is shown as SEQ ID NO:12.

SEQ ID NO:12>sp|P21741|MK_HUMAN Midkine OS=Homo sapiens OX=9606 GN=MDK PE=1
 SV=1
 MQHRGFLLLTLLALLALTSAVAKKKDKVKKGGPGSECAEAWAGPCTPSSKDCGVGFREGT
 CGAQTQIRICRVPCNWKKEFGADCKYKFENWGACDGGTGTKVRQGTLLKKARYNAQCQETI
 RVTKPCTPKTKAKAKAKKGGKGD

As used herein, the term “**MFGM**” has its general meaning in the art and refers to the
 lactadherin. An exemplary amino acid sequence for MFGM is shown as SEQ ID NO:13.

SEQ ID NO:13>sp|Q08431|MFGM_HUMAN Lactadherin OS=Homo sapiens OX=9606
 GN=MFGE8 PE=1 SV=3
 MPRPRLAALCGALLCAPSLLVALDICSKNPCHNGGLCEEISQEVRGDVFPSTCTCLKG
 YAGNHCETKCEPLGLENGNIANSQIAASSVRVTFLGLQHWVPELARNRAGMVNAWTPS
 SNDDNPWIQVNLRLRMWVTGVVTQASRLASHEYLKAFKVAYSLNGHEFDFIHDVNKKHK
 EFVGNWNKNVHVNLFETPVEAQYVRLYPTSCHTACTLRFELLGCELNGCANPLGLKNNS
 IPDKQITASSSYKTWGLHLFSWNPYSARLDKQGNFNAWVAGSYGNDQWLQVDLGSSKEVT
 GIITQGARNFGSVQFVASVKVAYSNDANWTEYQDPRTGSSKIFPGNWDNHSKKNLFT
 PILARYVRILPVAWHNRILRLLELLGC

As used herein, the term “**MMP-12**” has its general meaning in the art and refers to the
 macrophage metalloelastase. An exemplary amino acid sequence for MMP-12 is shown as SEQ
 ID NO:14.

SEQ ID NO:14>sp|P39900|MMP12_HUMAN Macrophage metalloelastase OS=Homo
 sapiens OX=9606 GN=MMP12 PE=1 SV=1
 MKFLILILLQATASGALPLNSSTSLEKNNVLFGERYLEKFYGLEINKLPVTKMKYSGNLM
 KEKIQEMQHFLGLKVTGQLDTSTLEMMHAPRCGVDPDVHVFREMPGGPVWRKHYITYRINN
 YTPDMNREDVDYAIRKAFQVWSNVTPKFSKINTGMADILVVFARGAHGDFHAFDGKGGI
 LAHAFPGSGIGGDAHFDEDEFWTTTHSGGTNLFLTAVHEIGHSLGLGHSSDPKAVMFPTY
 KYVDINTFRLSADDIRGIQSLYGDPEKQRLPNPDNSEPALCDPNLSFDAVTTVGNKIFF
 FKDRFFWLKVSRPKTSVNLISLWPTLPSGIEAAYEIEARNQVFLFKDDKYWLISNLRP
 EPNPKSIHSFGFPNFVKKIDAAVFNPRFYRTYFFVDNQYWRDERRQMMDPGYPKLITK
 NFQIGIPKIDAVFYSKNKKYFFQGSNQFEYDFLLQRLTKLKSNSWFGC

As used herein, the term “**Renin**” has its general meaning in the art and refers to the renin
 receptor. . An exemplary amino acid sequence for Renin is shown as SEQ ID NO:15.

SEQ ID NO:15>sp|O75787|RENH_HUMAN Renin receptor OS=Homo sapiens OX=9606
 GN=ATP6AP2 PE=1 SV=2
 MAVFVLLALVAGVLGNEFSILKSPGSVVFNRNGNWPIPERIPDVAALSMGFSVKEDLSW
 PGLAVGNLFHRPRATVMVMVKGVNKLALPPGSVISYPLENAVFPFSLDSVANSIHSLFSEE
 5 TPVVLQLAPSEERVYVMVGKANSVFEDLSVTLRQLRNRLFQENSVLSSPLNLSLRNNEVD
 LLFLSELQVLHDISSLLSRHKHLAKDHSPDLYSLELAGLDEIGKRYGEDSEQFRDASKIL
 VDALQKFADDMYSLYGGNAVVELVTVKSFDTSILIRKTRTILEAKQAKNPASPYNLAYKYN
 FEYSVVFNMVLWIMIALALAVIITSYNIWNMDPGYDSIIYRMTNQKIRMD

- 10 As used herein, the term “**high**” refers to a measure that is greater than normal, greater than a standard such as a predetermined reference value or a subgroup measure or that is relatively greater than another subgroup measure. For example, high levels of the protein of interest refers to a level of the protein that is greater than a normal level. A normal level may be determined according to any method available to one skilled in the art. High level of the protein may also
- 15 refer to a level that is equal to or greater than a predetermined reference value, such as a predetermined cutoff. High level of the protein may also refer to a level of the protein wherein a high protein subgroup has relatively greater levels of the protein than another subgroup. For example, without limitation, according to the present specification, two distinct patient subgroups can be created by dividing samples around a mathematically determined point, such
- 20 as, without limitation, a median, thus creating a subgroup whose measure is high (i.e., higher than the median) and another subgroup whose measure is low. In some cases, a “high” level may comprise a range of level that is very high and a range of level that is “moderately high” where moderately high is a level that is greater than normal, but less than “very high”.
- 25 As used herein, the term “**low**” refers to a level that is less than normal, less than a standard such as a predetermined reference value or a subgroup measure that is relatively less than another subgroup level. For example, low level of the protein means a level of the protein that is less than a normal level of in a particular set of samples of patients. A normal level of protein measure may be determined according to any method available to one skilled in the art. Low
- 30 level of the protein may also mean a level that is less than a predetermined reference value, such as a predetermined cutoff. Low level of the protein may also mean a level wherein a low level protein subgroup is relatively lower than another subgroup. For example, without limitation, according to the present specification, two distinct patient subgroups can be created by dividing samples around a mathematically determined point, such as, without limitation, a
- 35 median, thus creating a group whose measure is low (i.e., less than the median) with respect to another group whose measure is high (i.e., greater than the median).

As used herein, the term "**classification algorithm**" has its general meaning in the art and refers to classification and regression tree methods and multivariate classification well known in the art such as described in US 8,126,690 and WO2008/156617.

5 As used herein, the term "**support vector machine (SVM)**" is a universal learning machine useful for pattern recognition, whose decision surface is parameterized by a set of support vectors and a set of corresponding weights, refers to a method of not separately processing, but simultaneously processing a plurality of variables. Thus, the support vector machine is useful as a statistical tool for classification. The support vector machine non-linearly maps its n-
10 dimensional input space into a high dimensional feature space, and presents an optimal interface (optimal parting plane) between features. The support vector machine comprises two phases: a training phase and a testing phase. In the training phase, support vectors are produced, while estimation is performed according to a specific rule in the testing phase. In general, SVMs provide a model for use in classifying each of n subjects to two or more disease categories based
15 on one k-dimensional vector (called a k-tuple) of biomarker measurements per subject. An SVM first transforms the k-tuples using a kernel function into a space of equal or higher dimension. The kernel function projects the data into a space where the categories can be better separated using hyperplanes than would be possible in the original data space. To determine the hyperplanes with which to discriminate between categories, a set of support vectors, which lie
20 closest to the boundary between the disease categories, may be chosen. A hyperplane is then selected by known SVM techniques such that the distance between the support vectors and the hyperplane is maximal within the bounds of a cost function that penalizes incorrect predictions. This hyperplane is the one which optimally separates the data in terms of prediction (Vapnik, 1998 Statistical Learning Theory. New York: Wiley). Any new observation is then classified
25 as belonging to any one of the categories of interest, based where the observation lies in relation to the hyperplane. When more than two categories are considered, the process is carried out pairwise for all of the categories and those results combined to create a rule to discriminate between all the categories.

30 As used herein, the term "**Random Forests algorithm**" or "**RF**" has its general meaning in the art and refers to classification algorithm such as described in US 8,126,690 or WO2008/156617. Random Forest is a decision-tree-based classifier that is constructed using an algorithm originally developed by Leo Breiman (Breiman L, "Random forests," Machine Learning 2001, 45:5-32). The classifier uses a large number of individual decision trees and decides the class

by choosing the mode of the classes as determined by the subject trees. The subject trees are constructed using the following algorithm: (1) Assume that the number of cases in the training set is N, and that the number of variables in the classifier is M; (2) Select the number of input variables that will be used to determine the decision at a node of the tree; this number, m should be much less than M; (3) Choose a training set by choosing N samples from the training set with replacement; (4) For each node of the tree randomly select m of the M variables on which to base the decision at that node; (5) Calculate the best split based on these m variables in the training set.

10 **Methods of the present invention:**

Accordingly, the first object of the present invention relates to a method of predicting the clinical outcome of a patient suffering from chronic obstructive pulmonary disease (COPD) comprising the steps consisting of determining the level of a least one protein selected from the group consisting of BTK, Cyclophilin F, Tropomyosin 4, FGF-16, AREG, SHC1, 14-3-3 protein β/α , eIF-4H, CD40 ligand, H2A3, Midkine, MFGM, MMP-12, Renin and Carbonic anhydrase XIII wherein said level indicates the clinical outcome.

In some embodiments, the level of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15 proteins is determined in the blood sample.

Typically, the level of the protein is determined by an immunoassay. Such assays include, for example, competition assays, direct reaction assays sandwich-type assays and immunoassays (e.g. ELISA). The assays may be quantitative or qualitative. There are a number of different conventional assays for detecting formation of an immunocomplex. For example, the detecting step can comprise performing an ELISA assay, performing a lateral flow immunoassay, performing an agglutination assay, analyzing the sample in an analytical rotor, or analyzing the sample with an electrochemical, optical, or opto-electronic sensor. These different assays are well-known to those skilled in the art. In some embodiments, the devices are useful for performing an immunoassay according to the present invention. For example, in some embodiments, the device is a lateral flow immunoassay device. In some embodiments, the device is an analytical rotor. In some embodiments, the device is a dot blot. In some embodiments, the device is a tube or a well, e.g., in a plate suitable for an ELISA assay. In some embodiments, the device is an electrochemical sensor, an optical sensor, or an opto-electronic

sensor. The presence and amount of the immunocomplex may be detected by methods known in the art, including label- based and label-free detection. For example, label-based detection methods include addition of a secondary antibody that is coupled to an indicator reagent comprising a signal generating compound. The secondary antibody may be an anti-human IgG antibody. Indicator reagents include chromogenic agents, catalysts such as enzyme conjugates, fluorescent compounds such as fluorescein and rhodamine, chemiluminescent compounds such as dioxetanes, acridiniums, phenanthridiniums, ruthenium, and luminol, radioactive elements, direct visual labels, as well as cofactors, inhibitors and magnetic particles. Examples of enzyme conjugates include alkaline phosphatase, horseradish peroxidase and beta-galactosidase. Methods of label-free detection include surface plasmon resonance, carbon nanotubes and nanowires, and interferometry. Label- based and label-free detection methods are known in the art and disclosed, for example, by Hall et al. (2007) and by Ray et al. (2010) *Proteomics* 10:731-748. Detection may be accomplished by scanning methods known in the art and appropriate for the label used, and associated analytical software. In some embodiments, fluorescence labeling and detection methods are used to detect the immunocomplexes. A particularly useful assay format is a lateral flow immunoassay format. Antibodies to human or animal (e.g., dog, mouse, deer, etc.) immunoglobulins, or staph A or G protein antibodies, can be labeled with a signal generator or reporter (e.g., colloidal gold) that is dried and placed on a glass fiber pad (sample application pad or conjugate pad). Another assay is an enzyme linked immunosorbent assay, i.e., an ELISA. Typically, in an ELISA, the proteins are adsorbed to the surface of a microtiter well directly or through a capture matrix (e.g., an antibody). Residual, non-specific protein-binding sites on the surface are then blocked with an appropriate agent, such as bovine serum albumin (BSA), heat-inactivated normal goat serum (NGS), or BLOTTO (a buffered solution of nonfat dry milk which also contains a preservative, salts, and an antifoaming agent). The well is then incubated with the sample. The sample can be applied neat, or more often it can be diluted, usually in a buffered solution which contains a small amount (0.1-5.0% by weight) of protein, such as BSA, NGS, or BLOTTO. After incubating for a sufficient length of time to allow specific binding to occur, the well is washed to remove unbound protein and then incubated with an optimal concentration of an appropriate anti-immunoglobulin antibody (e.g., for human subjects, an anti-human immunoglobulin (α Hulg) from another animal, such as dog, mouse, cow, etc. that is conjugated to an enzyme or other label by standard procedures and is dissolved in blocking buffer. The label can be chosen from a variety of enzymes, including horseradish peroxidase (HRP), beta-galactosidase, alkaline phosphatase, glucose oxidase, etc. Sufficient time is allowed for specific binding to occur again, then the well is washed again to

remove unbound conjugate, and a suitable substrate for the enzyme is added. Color is allowed to develop and the optical density of the contents of the well is determined visually or instrumentally (measured at an appropriate wave length).

5 In some embodiments, multiplexing quantification is carried-out. In particular, determining the level of proteins in the samples may use one or more affinity reagents specific for the defined panel of proteins; physically contacting the panel with the sample identifying and quantitating the proteins that bind to the panel. A variety of different assays can be used to quantify the levels of such proteins. Multiplex arrays in several different formats based on the utilization of, for example, flow cytometry, chemiluminescence or electron-chemiluminescence technology, are well known in the art. Flow cytometric multiplex arrays, also known as bead-based multiplex arrays, include the Cytometric Bead Array (CBA) system from BD Biosciences (Bedford, Mass.) and multi-analyte profiling (xMAP®) technology from Luminex Corp. (Austin, Tex.), both of which employ bead sets which are distinguishable by flow cytometry. Each bead set is coated with a specific capture antibody. Fluorescence or streptavidin-labelled detection antibodies bind to specific capture antibody-biomarker complexes formed on the bead set. Multiple biomarkers can be recognized and measured by differences in the bead sets, with chromogenic or fluorogenic emissions being detected using flow cytometric analysis. In some embodiments, a multiplex ELISA coat multiple specific capture antibodies at multiple spots (one antibody at one spot) in the same well on a 96-well microtiter plate. Chemiluminescence technology is then used to detect multiple biomarkers at the corresponding spots on the plate. Determining the level of the proteins described herein can also be conducted using ProcartaPlex® Multiplex Immunoassays purchasable from Thermofischer. ProcartaPlex Multiplex Immunoassays are bead-based assays for protein quantification based on the principles of a sandwich ELISA with the use of Luminex® xMAP® (multi-analyte profiling) technology.

In some embodiment, high levels of: AREG, FGF-16, SHC1, 14-3-3 protein β/α , eIF-4H, Tropomyosin 4, Cyclophilin F, Carbonic anhydrase XIII, H2A3, CD40 ligand, BTK and low levels of: midkine, Lactadherin, MMP-12, renin indicate a low risk of worsening and/or high levels of: Midkine, Lactadherin, MMP-12, renin and low levels of: AREG, FGF-16, SHC1, 14-3-3 protein β/α , eIF-4H, Tropomyosin 4, Cyclophilin F, Carbonic anhydrase XIII, H2A3, CD40 ligand, BTK indicate a high risk of worsening.

In some embodiments, the method of the present invention further comprises comparing the level of protein with a predetermined reference value wherein detecting a difference between the level of the protein and the predetermined reference value indicates the clinical outcome (e.g. the risk of worsening). Typically, when the levels of protein are higher than the predetermined reference value, it is concluded that the patient has a low risk of worsening and conversely, when the levels of protein are lower than the predetermined reference value, it is concluded that the patient has a high risk of worsening.

In some embodiments, the predetermined reference value is a relative to a number or value derived from population studies, including without limitation, subjects of the same or similar age range, subjects in the same or similar ethnic group, and subjects having the same severity of COPD (e.g. the same GOLD stage). Such predetermined reference values can be derived from statistical analyses and/or risk prediction data of populations obtained from mathematical algorithms and computed indices. In some embodiments, retrospective measurement of the level of the protein in properly banked historical subject samples may be used in establishing these predetermined reference values. Accordingly, in some embodiments, the predetermined reference value is a threshold value or a cut-off value. The threshold value has to be determined in order to obtain the optimal sensitivity and specificity according to the function of the test and the benefit/risk balance (clinical consequences of false positive and false negative). Typically, the optimal sensitivity and specificity (and so the threshold value) can be determined using a Receiver Operating Characteristic (ROC) curve based on experimental data. For example, after determining the level of the protein in a group of reference, one can use algorithmic analysis for the statistic treatment of the measured levels of the protein in samples to be tested, and thus obtain a classification standard having significance for sample classification. The full name of ROC curve is receiver operator characteristic curve, which is also known as receiver operation characteristic curve. It is mainly used for clinical biochemical diagnostic tests. ROC curve is a comprehensive indicator that reflects the continuous variables of true positive rate (sensitivity) and false positive rate (1-specificity). It reveals the relationship between sensitivity and specificity with the image composition method. A series of different cut-off values (thresholds or critical values, boundary values between normal and abnormal results of diagnostic test) are set as continuous variables to calculate a series of sensitivity and specificity values. Then sensitivity is used as the vertical coordinate and specificity is used as the horizontal coordinate to draw a curve. The higher the area under the curve (AUC), the higher the accuracy of diagnosis. On the ROC curve, the point closest to the far upper left of the coordinate diagram

is a critical point having both high sensitivity and high specificity values. The AUC value of the ROC curve is between 1.0 and 0.5. When $AUC > 0.5$, the diagnostic result gets better and better as AUC approaches 1. When AUC is between 0.5 and 0.7, the accuracy is low. When AUC is between 0.7 and 0.9, the accuracy is moderate. When AUC is higher than 0.9, the accuracy is quite high. This algorithmic method is preferably done with a computer. Existing software or systems in the art may be used for the drawing of the ROC curve, such as: MedCalc 9.2.0.1 medical statistical software, SPSS 9.0, ROCPOWER.SAS, DESIGNROC.FOR, MULTIREADER POWER.SAS, CREATE-ROC.SAS, GB STAT V10.0 (Dynamic Microsystems, Inc. Silver Spring, Md., USA), etc.

In some embodiments, the method of the present invention further comprises a step consisting of calculating a score, representing an estimation of the clinical outcome and/or risk of worsening. Typically, the score is based on the level of one or more protein(s) determined in the blood sample and may typically include another factor. Typically, other risk factors may include additional features such as age, gender, obesity, diabetes mellitus, current or former smoking, body mass index... In some embodiments, the other risk factor is the GOLD stage. Based the above input features obtained from the subject, an operator can calculate a numerical function of the above list of inputs by applying an algorithm. For instance, this numerical function may return a number, i.e. score (R), for instance between zero and one, where zero is the lowest possible risk indication and one is the highest. This numerical output may also be compared to a threshold (T) value between zero and one. If the risk score exceeds the threshold T, it is meant that the patient has a poor clinical outcome and/or a high risk of worsening and if the risk score is under the threshold T, it is meant that the patient has good clinical outcome and/or a low risk of worsening.

In some embodiments, the method of the invention thus comprises the use of an algorithm. In some embodiments, the algorithm is a classification algorithm typically selected from Multivariate Regression Analysis, Linear Discriminant Analysis (LDA), Topological Data Analysis (TDA), Neural Networks, Support Vector Machine (SVM) algorithm and Random Forest algorithm (RF). In some embodiments, the score is generated by a computer program.

In some embodiments, the method of the present invention thus comprises a) determining the level of one or more protein(s) in the sample obtained from the subject; b) implementing an

algorithm on data comprising the level of the protein(s) so as to obtain an algorithm output; c) determining the clinical outcome and/or the risk of worsening.

The invention will be further illustrated by the following figures and examples. However, these examples and figures should not be interpreted in any way as limiting the scope of the present invention.

FIGURES:

Figure 1: Study flow diagram. Stable COPD patients (n=241) were included in the COBRA cohort at Visit 1 and serum SOMAscan analyses were performed. Two COPD clusters, Cluster 1 (n=126) and Cluster 2 (n=115) with distinct protein expression patterns were identified through unsupervised hierarchical clustering. The differentially expressed proteins in Cluster 2 were submitted to Gene Ontology (GO) pathways analysis for further functional enrichment study, then median expression values were generated for each COPD cluster. Association studies were performed between COPD clusters or corresponding protein and clinical parameters. The evolution of the clinical outcomes and of proteomic profiles was assessed in 163 COPD patients out of the 241 having a second visit. Expression patterns associated with Cluster 1 (n=97) and Cluster 2 (n=66) were verified in COBRA cohort at Visit 2. Protein signatures representative of each COPD cluster were identified to monitor clinical patient stability at Visit 2. SOMAscan data obtained in the COBRA cohort were validated in a separate group of 47 COPD patients originating from the Melbourne Longitudinal COPD Cohort (MLCC) cohort. Serum samples from n=50 healthy subjects were included for comparisons.

Figure 2. Changes in COPD clustering and subtypes at Visit 2. Changes in the levels of the 15 selected proteins at visits 1 and 2 in COPD patients switching of from C1 to C2 and from C2 to C1. Data are expressed as median log2 signal intensity and they are medians (95% CI) of the number of patients in each group, as indicated in brackets on the top of the panel. On the left side, are shown the biological functions attributed to each protein. Comparisons were made by Friedman Rank test and bold denotes statistical significance.

EXAMPLE:

Methods

Study populations

Between March 2007 and November 2016, 241 stable COPD patients aged between 46 and 82 years were included in the COBRA cohort (17,18) (CPP Ile-de-France I Ethics Committee, n° 09-11962) and written informed consent was obtained before inclusion. Demographic and clinical variables are listed in **Table 3**. Peripheral blood was obtained to determine total and differential leukocyte counts and to obtain serum aliquots for the measurement of the levels of hemoglobin and of C reactive protein (CRP) and for SOMAscan analyses. The evolution of the clinical outcomes and of proteomic profiles was assessed in 163 COPD patients out of the 241 having a follow up visit 7.5 ± 6.6 months (mean $\pm$ SD) after inclusion (**Table 3**). SOMAscan data obtained in the COBRA cohort were validated in a separate series of 47 COPD patients originating from the Melbourne Longitudinal COPD Cohort (MLCC) cohort, that was established to study the aetiology of frequent exacerbators in a community setting (**Table 4**) (19). Serum samples from n=50 control healthy subjects were used for comparisons (see study flow diagram in **Figure 1**).

SOMAscan proteomic assay

Total of 1305 analytes were quantified in patient serum using the SOMAscan high throughput proteomic assay (SomaLogic, Boulder, CO) at National Jewish Health (Denver, CO) (20). The raw SOMAscan data were standardized according to manufacturer's instructions (<http://somalogic.com/wp-content/uploads/2017/06/SSM-071-Rev-0-Technical-Note-SOMAscan-Data-Standardization.pdf>) and log2 transformed for downstream analyses. Proteins were described in the manuscript using SOMAscan target names.

Data processing

Unsupervised clustering was performed using R and differential expression analyses between the two clusters were conducted using Limma package (21). Proteins with Fold change >1.5 and False discovery rate (FDR) <0.05 were defined as significant. Functional analyses of the differentially expressed proteins were performed according to GO. For two group comparisons, t test was performed for data with normal distribution and Wilcoxon Rank Sum test was performed when normal distribution cannot be assumed. Pearson and Spearman correlation was performed for data with normal and unnormal distribution, respectively. ANOVA or Kruskal-Wallis, followed by Fisher exact test and Friedman Rank test were performed for multi-group and binary data comparisons, respectively. Additional statistical analyses and plots, including volcano plots and heatmaps were generated in R, SAS/JMP 15.2.1 or GraphPad Prism 8 (see

online supplement for details). Data processing, bioinformatical approaches and statistical analyses are described in details in the online supplement and illustrated in **Figure 1**.

Results

Study populations

Comparison of GOLD stage distribution, smoking habits, respiratory function, and medication in COPD patients of the COBRA cohort were comparable in the 163 COPD patients having two visits (**Table 3**). In contrast, there was a significant reduction in the incidence of patients having exacerbations ($P=0.003$) at Visit 2, as compared to Visit 1 (**Table 3**). When compared to the 163 COPD patients with 2 visits at inclusion, the remaining 78 patients having only Visit 1 had less severe disease, as shown by their higher distribution into GOLD 1 group, values of % predicted post-bronchodilator FEV1 and FEV1/FVC, DLCO, unscheduled medical visits, OCS use, but similar onset and number of exacerbations and comorbidities (**Table 3**).

The 47 COPD patients from the MLCC cohort were more severe and symptomatic in comparison to those from the COBRA cohort, in terms of GOLD stage, airflow obstruction, incidence of cough, treatments with muscarinic antagonists, long-term inhaled steroids (ICS) and oxygen therapy (**Tables 3 and 4**).

Unsupervised hierarchical clustering at inclusion

Unsupervised hierarchical clustering led to the identification of two clusters of the 241 COPD subjects at inclusion (126 patients in Cluster 1 and 115 in Cluster 2) with distinct protein expression patterns. A total of 96 proteins showed significantly differential expression between the two clusters (fold change >1.5 and $FDR < 0.05$). Of these, 90 and 6 proteins showed significantly higher levels in Cluster 2 and in Cluster 1, respectively (**Data not shown** and **Table 1**). We generated a score for the median expression of the 90 proteins associated with Cluster 2 and confirmed a significant difference between the two clusters (**Data not shown**).

Association between biological pathways and COPD clusters

To characterize the biological processes and molecular function associated with each subtype of COPD subjects, we performed GO pathway analysis on the Top 90 up-regulated proteins in Cluster 2 ($FDR < 0.0001$). Our data highlighted the enrichment of this cluster in hallmarks related to lung immunity/host defense, cell fate/repair/remodelling and mitochondrial/metabolic activities (**Data not shown** and **Tables 1 and 5**).

Clinical characteristics of Cluster 1- and Cluster 2- COPD patients

Then, we examined whether the two subtypes of COPD exhibited distinct clinical characteristics by considering most of the parameters used to describe the COBRA cohort at inclusion (**Table 2**). Patients from Cluster 2 had a lower incidence and number of exacerbations and of unscheduled medical visits in the previous year, compared with those from Cluster 1 (**Table 2**). The proportion of COPD patients with emphysema, as assessed by CT scan and/or by the measurement of transfer factor of the lung for carbon monoxide (DLCO), was also lower in Cluster 2 than in Cluster 1 (**Table 2**). Further, Cluster 2 patients had a lower prevalence of hypertension, diabetes and obstructive sleep apnea (**Table 2**). In addition, the proportion of patients requiring long-acting β_2 -agonists (LABA) alone was significantly lower in Cluster 2 than in Cluster 1 and this was accompanied by a trend towards a lower rate of patients treated with LABA, in combination with long-lasting muscarinic antagonists (LAMA) (**Table 2**). In contrast, the incidence of ICS use, alone or in combination with LABA and LAMA, was not significantly different between the two clusters (**Table 2**).

Finally, the proportion of COPD patients necessitating anti-hypertensive drugs, statins, or other therapies was lower in Cluster 2 than in Cluster 1 (**Table 2**). No difference was observed between the two clusters in terms of GOLD stages, biology, respiratory function, symptoms and serum levels of CRP (*data not shown*).

Correlation between cluster-associated and key clinical parameters

Correlation analyses showed positive associations between lower incidence of exacerbations and/or prevalence of emphysema and circulating levels of different biomarkers implicated in the regulation of EGFR signalling pathways (eg. SHC1, AREG, GRB2 adapter protein), of host defense and innate immune responses (BTK), oxidant stress and hypoxia (eg. cyclophilin F, α -Synuclein and Carbonic anhydrase XIII, 14-3-3 protein β/α and ζ/δ , BAD), wound healing, migration and survival (eg. Tropomyosin 4, 14-3-3 protein β/α and ζ/δ , eIF-5A-1, FGF-16, PA2G4, TCTP, BAD), as well as some components of epithelial-mesenchymal remodelling (eg. RAC1, GRB2 adapter protein, ARG11, Prostatic binding protein, DRG-1, CPNE1). Positive correlations were also found between lower prevalence of emphysema and levels of biomarkers of T cell activation (CD40 ligand), metabolism (M2-PK and 6-phosphogluconate dehydrogenase) and proteostasis (eg. Ubiquitin+1, SUMO3, UBC9, Sorting nexin 4, SNAA, UFM1) (*Data not shown*).

Finally, a negative association was found between serum levels of MMP-12 and renin, which were up-regulated in Cluster 1, and lower prevalence of emphysema (*Data not shown*).

Investigation of Clusters 1 and 2 in an independent COPD cohort and in healthy subjects

We next validated these clusters in another COPD population consisting of 47 subjects (19). Unsupervised hierarchical clustering, with the top 10% proteins by expression variation, showed two clusters within these COPD subjects (34 and 13 subjects in Clusters 1 and 2, respectively) where 125 proteins were differentially expressed between Cluster 1 and Cluster 2 (Fold change > 1.5 and FDR < 0.05) (***Data not shown***). While 114 proteins were significantly up-regulated in Cluster 2, 71 of these (62%), overlapped with the proteins that were higher in Cluster 2 of the COBRA cohort at visit 1 (***Data not shown***). We also found no similarities between the 2 cohorts regarding the few down-regulated proteins in Cluster 2. We calculated the score (median expression) of the 90 Cluster 2-associated proteins defined in the COBRA cohort, and showed that it was significantly up-regulated in Cluster 2, as compared to Cluster 1 in the MLCC cohort (***Data not shown***).

Comparisons of the clinical parameters between the two clusters showed significant lower prevalence of emphysema and of incidence of hospitalizations for COPD during the last 12 months, as well as a diminished requirement of treatment with LABA in Cluster 2 compared to Cluster 1 (**Table 6**). This was accompanied by a non significant trend towards a lower onset of exacerbations and of cardio-metabolic co-morbidities (**Table 6**).

To determine whether the subtyping was specific of COPD, we performed the same analyses in serum samples from a group of 50 healthy donors. Although hierarchical clustering generated two subtypes within the healthy cohort, the differential expression of Cluster 2-associated proteins, as defined in the COBRA cohort, was very weak and it was not observed in these subjects (median log₂ signal intensity between Cluster 1 (n=19 subjects) and Cluster 2 (n=31 subjects) [95% CI]: 10.6 [10.5-10.7] and 10.7 [10.6-10.8], respectively, P=0.13).

Generation of protein signatures predictive of the newly identified COPD endotypes

To establish a short protein signature reflecting changes in clinical outcomes in relation with lung biological processes, we generated a fingerprint of 15 biologically relevant biomarkers composed of 10% of the top 96 differentially expressed proteins that were selected from the different GO biological processes identified (***Data not shown***). These biomarkers included 11 and 4 proteins representative of Cluster 2 and 1, respectively (***Data not shown***). Specifically, Cluster 2-associated Metabolism/Mitochondria markers, namely Cyclophilin F, Carbonic anhydrase XIII and H2A3, were sorted from GO pathways involved in mitochondrial membrane organization and permeability and cellular response to oxidative stress and to

environmental stimulus, respectively, whereas Immunity/Defense markers, including BTK and CD40 ligand, were sorted from activation of innate immune response and T cell regulation, respectively (**Data not shown**). To cover the different aspects of Cell fate/Repair/Remodeling processes, rational ranking of differentially expressed proteins led to the prioritization of: 1) the representative transcription factor, eIF-4H, that is involved in the post-translational regulation of gene expression; 2) the two growth factors, AREG and FGF-16, that promote tissue regeneration through ERBB signaling pathway and that positively regulate cell proliferation; 3) the cytosolic protein regulating cell proliferation, SHC1, 4) the key regulator of apoptotic and nutrient-sensing signaling, 14-3-3 protein β/α ; and 5) the tissue-remodeling mediator, Tropomyosin 4, that is involved in actin filament organisation. Lastly, to predict Cluster 1-COPD endotype, a selection of 4 of the 6 down-regulated proteins in Cluster 2 (**Data not shown**), was performed, after exclusion of C3b that failed to show any differential expression in the MLCC cohort. These 4 proteins included two hallmarks of tissue injury (MMP-12 and renin) and two modulators of lung immune responses (Midkine and Lactadherin - MFGM) (**Figure 2**).

Follow-up analysis at second visit

To determine whether the 2 clusters were maintained in the 163 subjects from COBRA cohort during a follow-up visit, we first performed unsupervised hierarchical clustering on these subjects using the top 10% proteins by expression variation. Consistent with the results at inclusion, we confirmed the existence of two distinct expression patterns at Visit 2 (97 subjects in Cluster 1 and 66 subjects in Cluster 2) (**Data not shown**). The expression of 86 proteins was significantly different between the two clusters (Fold change > 1.5 and FDR < 0.05), with 83 up-regulated and 3 down-regulated in Cluster 2 when compared to Cluster 1. Seventy-five out of these 83 upregulated proteins (eg. 90%) were also elevated in Cluster 2 at the initial visit, while one of the 3 down-regulated proteins in Cluster 2 (eg. Lactadherin) was also reduced in Cluster 2 at Visit 1. Furthermore, the protein score (median expression) of the 90 Cluster 2-associated proteins defined at Visit 1 was significantly higher in Cluster 2 at Visit 2, than at Visit 1 (**Data not shown**).

Next, we demonstrated that clustering pattern was significantly consistent between the two visits, with 108 subjects among 163 showing the same cluster identify (68 for Cluster 1 and 40 for Cluster 2) ($P < 0.001$, Fisher exact test). However, 55 subjects changed their cluster identity (29 switched from Cluster 1 at Visit 1 to Cluster 2 at Visit 2, 26 switched from Cluster 2 at Visit 1 to Cluster 1 at Visit 2) (**Data not shown**). COPD patient switching from to Cluster 1 to 2

displayed lower incidence of exacerbations, of unscheduled medical visits and of hospitalizations for COPD in the previous year. In contrast, COPD patients switching from Cluster 2 to 1 showed a significant higher incidence of hospitalizations for COPD and of LAMA use (*Data not shown*).

5 By monitoring the expression pattern of the 15 biomarkers of the protein signature in the 163 patients that attended both Visits 1 and 2, we were able to map the 2 groups of COPD patients (n=55) that exhibited cluster switching, eg. Cluster 2 to 1, or the opposite, while the remaining 108 patients showed similar expression pattern at both visits (*Figure 2*). The levels of AREG, FGF-16, SHC1, 14-3-3 protein β/α , eIF-4H, Tropomyosin 4, Cyclophilin F, Carbonic
10 anhydrase XIII, H2A3, and BTK were upregulated, whereas those of lactadherin and MMP-12 were reduced in COPD patients switching from Cluster 1 to Cluster 2 at Visit 2 (*Data not shown and Figure 2*). Mirroring these changes, the 11 proteins associated with Cluster 2 were significantly down-regulated in COPD patients switching from Cluster 2 to Cluster 1 at Visit 2, whereas those of Midkine, Lactadherin, MMP-12 and renin were increased (*Data not shown*
15 *and Figure 2*).

Discussion

Using SOMAscan, we profiled the serum levels of 1305 proteins in 241 patients with COPD from the COBRA cohort (18) and identified two distinct clusters defined by the differential
20 expression of 96 proteins. These two clusters appeared to be clinically relevant since Cluster 2 showed lower number and incidence of acute exacerbations, unscheduled medical visits, hospitalizations for COPD within the previous year and reduced prevalence of emphysema than Cluster 1.

Lung function was not different between the 2 Clusters, which could be explained by the
25 elevated proportion of patients of GOLD stages 1 and 2 in the COBRA cohort.

The occurrence of co-morbidities, particularly hypertension, diabetes and obstructive sleep apnoea, the need of LABA and of treatments for cardiovascular and metabolic comorbidities (i.e., anti-hypertensive drugs and statins), were also lower in Cluster 2 than in Cluster 1.

We also confirmed the presence of these 2 patient endotypes with a significant conservation of
30 cluster-associated proteins in 47 COPD patients of the MLCC cohort (19). Although we acknowledge that this low number of patients represents a study limitation, we observed a reduction in the prevalence of emphysema and on the incidence of hospitalizations for COPD during the last 12 months in Cluster 2, compared to Cluster 1 and a trend towards a lower occurrence of exacerbations and of cardio-metabolic co-morbidities. These data suggest that

the SOMAscan technique offers a powerful and largely reproducible screening tool to identify specific serum biomarkers predictive of new COPD-related endotypes.

Through computational annotation, combined to protein mapping into GO pathways of the Cluster-associating proteins in the COBRA cohort, we showed that serum samples of patients in Cluster 2 were significantly enriched in hallmarks of cell fate/remodelling/repair, proteostasis (metabolism and mitochondria) and lung immunity/defense. This was accompanied by an upregulation of biological processes involved in epithelial/mesenchymal/vascular remodelling, pro-survival signals and cell renewal activity, and in metabolic signalings including lung hypoxic, metabolic and mitochondrial pathways.

These changes may reflect high protein turnover and glycolytic activity, mitochondrial functions and proteostasis that are aimed at maintaining mucosal barrier under oxidative stress during lung remodelling in COPD (22). Consistently, previous data reported a rise in proteins involved in glycolysis/gluconeogenesis pathways and in folding/aggregation processes in patients with COPD (23,24). This enhanced glycolytic activity may regulate metabolic energy in proliferating immune and structural cells upon tissue remodelling under oxidative stress.

To distinguish between the different COPD clusters, we refined the top 96 differentially expressed proteins into a clinically feasible fingerprint composed of 15 biologically relevant biomarkers, that included 4 and 11 up-regulated proteins in Cluster 1 and Cluster 2, respectively.

The 11 up-regulated biomarkers included the pro-survival protein, 14-3-3 protein β/α , the transcription factor regulating stem cell renewal, eIF4H, and the growth factors promoting stem cell activity, AREG, SHC1 and FGF-16. These factors have been reported to promote tissue repair (25–28), suggesting that Cluster 2-COPD patients exhibit a better lung regeneration than those in Cluster 1. The mesenchymal biomarker, Tropomyosin 4, was also augmented in Cluster 2 patients. This protein was initially documented for its actin-binding properties in muscle cells and its expression increased in muscle fibres of diaphragm of COPD patients upon adaptation aimed at increasing resistance to fatigue (29). Recently, a role of Tropomyosin 4 in the regulation of adhesion, division, motility and apoptosis in non-muscular cells was also suggested (30).

Carbonic anhydrase XIII, that was also elevated in Cluster 2 patients, is a protein responsible for carbon dioxide hydration and, therefore, it is considered as one of the main hallmarks of systemic and local oxidative stress and nitration. The expression of carbonic anhydrase XIII was shown to be decreased in skeletal muscles of COPD patients and it was inversely correlated with loss of force (31). Together, these findings suggest that elevated levels of Tropomyosin 4

and of Carbonic anhydrase XIII in Cluster 2 COPD patients translate a better efficiency of the antioxidant systems to improve muscle dysfunction in response to chronic exercise (32).

Representative hallmarks of proteostasis (eg. Cyclophilin F, Carbonic anhydrase XIII and H2A3) , and of innate immune response, host defense and inflammasome activity (BTK and

CD40 ligand) (32–35) were also more elevated in Cluster 2, than in Cluster 1 COPD patients.

Cyclophilin F is a ubiquitously expressed protein belonging to the immunophilin family, that has been involved in protein folding, trafficking, T-cell activation and mitochondrial permeability (36). Cyclophilins are secreted under inflammatory stimuli and oxidative stress, including in COPD (37). Elevation of these proteins, together with that of BTK and CD40 ligand, suggest an improved immune and inflammatory responses upon lung injury and oxidative stress in Cluster 2-COPD patients.

Finally, among the 4 proteins of our fingerprint that were upregulated in Cluster 1, Midkine and, Lactadherin (MFGM), have been shown to be elevated during several chronic immune and metabolic disorders, such as atherosclerosis, cardiac, kidney and metabolic diseases (38–40).

Elevated levels of these proteins are consistent with higher incidence of hypertension, diabetes and use of anti-hypertensive drugs and statins and they support the contribution of cardiometabolic co-morbidities in COPD severity (6). The levels of MMP-12 were also higher in Cluster 1 than in Cluster 2 patients and this was associated with a greater prevalence of emphysema. This finding is consistent with the crucial role played by this protease in lung injury secondary to degradation of various extracellular matrix components seen in COPD (41). Likewise, elevated renin, an enzyme involved in the renin-angiotensin II-aldosterone axis that regulates blood pressure, was associated with higher prevalence of hypertension in Cluster 2 patients (42).

Profiling serum protein at Visit 2 showed that approximately 34% of COPD patients switched their clusters. This switch was not linked to acute events (eg. exacerbations, pneumonia), or to co-morbidities, all patients being in a stable state during the month preceding blood sampling. Patients switching from Cluster 1 to Cluster 2 (18%) showed significant lower incidence of exacerbations, of unscheduled medical visits and of hospitalizations for COPD, whereas switching from Cluster 2 to Cluster 1 (16%) was associated with an increase in the onset of hospitalizations for COPD, with a trend towards a higher prevalence of emphysema and with a lower incidence of treatment with LAMA. Inadequate LAMA use may explain, at least in part, a lower beneficial clinical response observed in these patients.

Importantly, the 15-protein signature allowed the monitoring of COPD subjects that switched their clusters at a follow-up visit. Thus, patients switching from Cluster 1 to Cluster 2 had higher

levels of 10 out of the 11 Cluster 2-associated proteins that are involved in cell fate, remodeling and repair (AREG, FGF-16, SHC1, 14-3-3 protein β/α , eIF-4H and Tropomyosin 4), in those regulating metabolism and mitochondrial functions (cyclophilin F and carbonic anhydrase XII), and in BTK CD40 ligand, that participate in immune responses and host defense. These results mirrored those obtained in patients switching from Cluster 2 to Cluster 1, where the levels of most of these proteins were downregulated at Visit 2, compared to Visit 1, while Cluster 1-associated biomarkers (Midkine, Lactadherin -MFGM-, MMP-12 and renin) were upregulated. These findings suggest the existence of an imbalance of immune responses associated with increased tissue damage in these patients.

Overall, this study identified a signature of 15 proteins that characterized two different COPD phenotypes and distinguished clinically stable from unstable patients. These proteins are major contributors of tissue injury, remodeling and repair, immunity and defense, and skeletal muscle dysfunction, suggesting that enhanced tissue renewal and remodeling, in compensation of tissue degradation mechanisms, and an improvement of immune and host defense responses to foreign insults, particularly tobacco smoke, contributed to a better disease outcome.

Of significance, FGF-16, a growth factor contributing to tissue regeneration (25,26), may represent a starting point for the development of new therapeutics. Consistently, the airway administration of FGF-2, another member of FGF family promoting tissue repair, was reported to reduce emphysema and to enhance lung repair in vivo, possibly by attenuating inflammation and alveolar cell death (43). In addition, randomised clinical trials performed in patients with periodontitis and osteoarthritis, two chronic inflammatory diseases with progressive tissue degeneration, demonstrated improved tissue repair upon the administration of recombinant human FGF-2 (44,45). Whether FGF-16 promotes similar beneficial effects in COPD patients deserves further investigation.

In conclusion, the clinical use of this signature would allow a better management of COPD patients in terms of monitoring, treatment options, selection for inclusion in clinical trials and for evaluating novel therapies.

TABLES:

Table 1: Biological classification of the top 90 proteins enriched in Cluster 2 patients

Biological processes (originating from the GO database)	Proteins
ERBB signalling pathway	AREG , GRB2 adapter protein, SRCN1, PDPK1, PKC-A, SHC1 , MK01

- Wound healing
 PKC-B-II, VAV, SRCN1, LYN, SMAD2, LYNB, Caspase-3, PDPK1, RAC1, Haemoglobin, METAP1, PRKACA, annexin I, PKC-A, FYN, 14-3-3 protein ζ/δ , PTP-1C, GPVI, ERK-1, NCC27, MK01
- Posttranscriptional regulation of gene expression
eIF-4H, 14-3-3 protein ζ/δ , 14-3-3 protein ζ/δ , GAPDH, PKC-A, RPS6KA3, hnRNP A2/B1, ERK-1, MK01, eIF-5A-1
 SRCN1, NDP kinase B, Caspase-3, PDPK1, CK2-A1:B, Sphingosine kinase 1, α -Synuclein, annexin I, PKC-A, RPS6KA3, UFM1, FYN, 14-3-3 protein ζ/δ , PPID, TCTP, PA2G4, **SHC1**, BAD, eIF-5A-1, **CD40 ligand**, Ubiquitin+1, Lactoferrin, Azurocidin, STAT3
- Negative regulation of programmed cell death
AREG, FGF16, SHC1, CK2-A1:B, PTP-1C, BAD, MK01, eIF-5A-1, STAT3
- Positive regulation of cell proliferation
BTK, LYN, SMAD2, LYNB, NDP kinase B, RAC1, annexin I, PKC-A, RPS6KA3, FYN, PA2G4, BAD, CPNE1, NCC27, eIF-5A-1, Lactoferrin
- Positive regulation of cell differentiation
BTK, PKC-B-II, SRCN1, LYN, SMAD2, LYNB, NDP kinase B, PDPK1, RAC1, Sphingosine kinase 1, annexin I, PKC-A, RPS6KA3, FYN, HXK2, PA2G4, BAD, CPNE1, NCC27, eIF-5A-1, Lactoferrin, STAT3
- Positive regulation of developmental process
 FER, PKC-B-II, IMB1, SRCN1, NSF1C, SBDS, CK2-A1:B, IF4G2, PRKACA, PKC-A, RPS6KA3, RAN, 41, PTP-1C, PA2G4, UBC9, ERK-1, MK01
- Cell cycle
BTK, FER, LYN, LYNB, PDPK1, ARG11, Sorting nexin 4
- Regulation of myeloid leukocyte mediated immunity
BTK, FER, VAV, GRB2 adapter protein, LYN, LYNB, PDPK1, RAC1, **SHC1**, ERK-1, MK01
- Fc- ϵ receptor signalling pathway
BTK, SP-D, SRCN1, LYN, LYNB, PDPK1, UBE2N, PRKACA, RPS6KA3, FYN, Ubiquitin+1
- Activation of innate immune response
CD40 ligand, CSK, SRCN1, Caspase-3, PDPK1, PTP-1C, DUSP3
- Regulation of T cell activation

- Positive regulation of defense response
 - Neutrophil degranulation
 - ADP metabolic process
 - Mitochondrial membrane organization and permeability
 - Cellular response to oxidative stress
 - Positive regulation of protein modification process
 - Cellular response to environmental stimulus
 - Actin filament organisation
- BTK**, SP-D, VAV, SRCN1, LYN, LYNB, PDPK1, UBE2N, α -Synuclein, PRKACA, RPS6KA3, FYN, ARG11, ERK-1, Ubiquitin+1, Lactoferrin, Sorting nexin 4, IMB1, NDP kinase B, CK2-A1:B, RAC1, Haemoglobin, IMDH1, PTP-1C, PA2G4, ARG11, Cyclophilin A, CPNE1, BPI, MK01, Lactoferrin, Azurocidin, GAPDH, Myokinase, HXK2, Triosephosphate isomerase, BAD, **Cyclophilin F**, α -Synuclein, 14-3-3 protein ζ/δ , HXK2, BAD, STAT3, **14-3-3 protein ζ/δ** , **Carbonic anhydrase XIII**, FER, SRCN1, NDP kinase B, α -Synuclein, annexin I, ARG11, ERK-1, MK01, **AREG**, CSK, SRCN1, LYN, LYNB, PDPK1, CK2-A1:B, RAC1, Sphingosine kinase 1, UBE2N, α -Synuclein, PRKACA, PKC-A, FYN, UBC9, EDAR, **SHC1**, ERK-1, MK01, **CD40 ligand**, Ubiquitin+1, Lactoferrin, Azurocidin, STAT3, Ubiquitin+1, **H2A3**, GRB2 adapter protein, **Tropomyosin 4**, GRB2 adapter protein, FER

Abbreviations: 41: Protein 4.1; AREG: Amphiregulin; ARG11: Arginase-1; BAD: Bcl2-associated agonist of cell death; BPI: Bactericidal permeability-increasing protein; BTK: Tyrosine-protein kinase BTK; CK2-A1:B: Casein kinase II 2-alpha:2-beta heterotetramer; CPNE1: Copine1; CSK: Tyrosine-protein kinase CSK; PPID: Peptidyl-prolyl cis-trans isomerase D or Cyclophilin D; Cyclophilin F; DUS3: Dual specificity protein phosphatase 3; EDAR: Tumor necrosis factor receptor superfamily member EDAR; eIF-4H: Eukaryotic translation initiation factor 4H; eIF-5A-1: Eukaryotic translation initiation factor 5A-1; ERK-1: Mitogen-activated protein kinase 3 (MAPK3); FER: Tyrosine-protein kinase Fer; FGF-16: Fibroblast growth factor 16; FYN: Tyrosine-protein kinase Fyn; GAPDH: Glyceraldehyde-3-phosphate dehydrogenase; GPVI: Platelet glycoprotein VI; GRB2 adapter protein: Growth factor receptor-bound protein 2; hnRNP A2/B1: Heterogeneous nuclear ribonucleoproteins A2/B1; HXK2: Hexokinase-2; IF4G2: Eukaryotic translation initiation factor 4 gamma 2; IMB1: Importin subunit β -1; IMDH1: Inosine-5'-monophosphate dehydrogenase 1; LYN: Tyrosine-protein kinase Lyn; LYNB: Tyrosine-protein kinase Lyn, isoform B; MK01: Mitogen-activated protein kinase 1; M2-PK: Pyruvate kinase PKM; METAP1: Methionine aminopeptidase 1; NCC27 or CLIC1: Chloride intracellular channel protein 1; NDP kinase B: Nucleoside diphosphate kinase B; NSF1C: NSFL1 cofactor p47; PA2G4: Proliferation-associated protein 2G4; PDPK1: 3-phosphoinositide-dependent protein kinase 1; PRKACA: cAMP-dependent protein kinase catalytic subunit α ; PKC-A: Protein kinase C, α type; PKC-B-II: Protein kinase C, β type (splice variant β -II); PTP-1C or

PTPN6: Tyrosine-protein phosphatase non-receptor type 6; RAC1: Ras-related C3 botulinum toxin substrate 1; RAN: GTP-binding nuclear protein Ran; RPS6KA3: Ribosomal protein S6 kinase α 3; SBDS: Ribosome maturation protein SBDS; SHC1: SHC-transforming protein 1; SMAD2: Mothers against decapentaplegic homolog 2; SP-D: Pulmonary surfactant-associated protein D; SRCN1 or SRC: Proto-oncogene tyrosine-protein kinase Src; STAT3: Signal transducer and activator of transcription 3; SUMO3: Small ubiquitin-related modifier 3; TCTP: Translationally-controlled tumour protein; UBC9: SUMO-conjugating enzyme UBC9; UBE2N: Ubiquitin-conjugating enzyme E2 N; UFM1: Ubiquitin-fold modifier 1; VAV: Proto-oncogene vav.

In bold are indicated the proteins belonging to the short signature.

Table 2: Differences in clinical characteristics between COPD patients of Cluster 1 and Cluster 2 at Visit 1

Parameter	Cluster 1 - number *	Cluster 2 - number *	Values in Cluster 1	Values in Cluster 2	P value †
Body Mass Index (kg per m ²)	126	115	27.1 ± 6.8	25.8 ± 5.1	0.09
<i>Smoking history</i>					
• Never smokers – no. (%)	126	115	1 (1)	7 (6)	0.02
<i>Respiratory function and disease control</i>					
• With emphysema no. (%) ‡	126	115	65 (52)	27 (31)	<0.0001
• DLCO - %	42	73	54.4 ± 22.6	61.7 ± 20.9	0.08
• With exacerbations in the previous 12 months - no. (%)	126	115	77 (61)	51 (44)	0.01
• Number of exacerbations in the previous 12 months - no.	126	115	1.61 ± 0.25	1.24 ± 0.20	0.02
• With unscheduled medical visits in the previous 12 months - no. (%)	126	115	67 (53)	38 (33)	0.0002
• With hospitalizations for COPD in the previous 12 months – no. (%)	126	115	31 (25)	17 (15)	0.08
<i>Comorbidities</i>					
• Cardiovascular – no. (%)	126	115	61 (48)	42 (37)	0.07
• Hypertension – no. (%)	126	115	49 (39)	28 (24)	0.02
• Diabetes – no. (%)	126	115	21 (17)	7 (6)	0.009
• Obstructive sleep apnea – no. (%)	126	115	17 (13)	5 (4)	0.04
<i>Treatments</i>					
• On SABA – no. (%)	125	106	85 (68)	59 (56)	0.06
• On LABA alone – no. (%)	125	108	20 (16)	6 (6)	0.02
• On LAMA alone – no. (%)	125	108	9 (7)	7 (6)	0.83
• On ICS alone – no. (%)	125	108	1 (1)	3 (3)	0.25
• Daily dose of ICS alone - µg of equivalents beclomethasone	1	3	1000	517 ± 275	0.27
• On LABA + LAMA – no. (%)	125	108	15 (12)	7 (6)	0.06
• On LABA + LAMA + ICS – no. (%)	125	108	44 (35)	34 (31)	0.55
• On anti-hypertensive drugs – no. (%)	124	108	52 (42)	28 (26)	0.01
• On statins – no. (%)	124	108	37 (30)	19 (18)	0.03
• Other – no. (%)	124	107	66 (53)	36 (34)	0.003
• Adherence to treatment – no. (%)	124	100	107 (86)	96 (96)	0.02

Abbreviations: DLCO, transfer factor of the lung for carbon monoxide; ICS, inhaled corticosteroids; SABA, short-acting β2-agonists; LABA, long-acting β2-agonists; LAMA, long-lasting muscarinic antagonists.

* indicates the number of patients with each available variable

Data are expressed as numbers (%) and as means ± SD

† Students' t test or Fisher exact test 2-tailed

‡ Diagnosis of emphysema was established by CT scan and/or by the measurement of DLCO

Bold denotes statistical significance

Table 3: characteristics of COPD patients of the COBRA cohort

Parameter	n *	All patients (n = 244)		Patients with 1 visit (n=78)		Patients with 2 visits (n = 163)		P value †	P value ‡
		n *		n *		n *			
		Visit 1		Visit 1 only		Visit 1	Visit 2		
Male sex – no. (%)	241	164 (68.1)	78	50 (64.1)	163	114 (69.9)	-	0.37	-
Age (years)	241	63.1 ± 9.8	78	61.6 ± 11.4	163	63.8 ± 8.9	64.5 ± 9.0	0.14	-
Caucasian origin – no. (%)	239	224 (93.7)	76	70 (92.1)	163	154 (94.5)	-	0.49	-
Other origin – no. (%)	239	15 (6.3)	76	6 (7.9)	163	9 (5.5)	-	0.49	-
GOLD stages									
• GOLD I – no. (%)	238	54 (22.7)	75	24 (32.0)	159	28 (17.6)	28 (17.6)	0.03	0.99
• GOLD II – no. (%)	238	87 (36.5)	75	20 (26.6)	159	66 (41.5)	66 (41.5)	0.03	0.99
• GOLD III – no. (%)	238	53 (22.3)	75	14 (18.7)	159	38 (23.9)	36 (22.7)	0.03	0.99
• GOLD IV – no. (%)	238	44 (18.5)	75	17 (22.7)	159	27 (17.0)	29 (18.2)	0.03	0.99
Smoking history									
• Former smoker – no. (%)	241	149 (61.8)	78	41 (52.6)	163	108 (66.3)	108 (66.3)	0.002	1.00
• Packs per year in former smokers – no.	147	49.1 ± 23.9	39	49.7 ± 22.2	101	49.7 ± 25.1	49.9 ± 25.1	0.99	0.96
• Active smokers – no. (%)	241	84 (34.9)	78	30 (38.5)	163	54 (33.1)	54 (33.1)	0.002	1.00
• Packs per year in active smokers – no.	82	35.8 ± 21.7	28	35.1 ± 20.8	47	36.0 ± 23.8	34.9 ± 21.9	0.89	0.83
Body Mass Index (kg per m ²)	240	26.5 ± 6.1	77	25.7 ± 5.9	163	26.8 ± 6.2	-	0.21	-
Biology									
• Blood leucocytes (no. per mm ³)	165	7500 (6500 - 8900)	43	7400 (6400 - 10000)	96	7350 (6500 - 8750)	7100 (6150 - 8150)	0.37	0.34
• Blood eosinophils (no. per mm ³)	165	162 (104 - 230)	43	146 (94 - 213)	96	165 (103 - 225)	162 (117 - 244)	0.37	0.67
• With blood eosinophils ≥ 300 per mm ³ – no. (%)	165	27 (16.4)	43	4 (9.3)	96	17 (17.7)	12 (12.5)	0.21	0.32
• Blood neutrophils (no. per mm ³)	163	4752 (3840 - 5950)	42	4479 (3782 - 6486)	95	4680 (3835 - 5950)	4560 (3843 - 5265)	0.74	0.39
• Blood lymphocytes (no. per mm ³)	165	1830 (1406 - 2280)	43	1920 (1539 - 2438)	96	1648 (1302 - 2190)	1782 (1387 - 2184)	0.04	0.49
• Blood monocytes (no. per mm ³)	161	560 (444 - 720)	40	603 (469 - 740)	96	532 (404 - 698)	555 (463 - 667)	0.03	0.25
• Hemoglobin – g per deciliter	115	14.4 ± 1.4	32	14.1 ± 1.6	72	14.6 ± 1.3	14.3 ± 1.4	0.10	0.25
• CRP – mg per Liter	112	7.6 ± 7.6	22	6.5 ± 4.6	73	7.8 ± 8.1	6.8 ± 7.5	0.35	0.47
With CRP ≥ 3 mg per Liter – no. (%)	112	85 (75.9)	22	17 (77.3)	73	54 (74.0)	54 (74.0)	0.76	1.00
Blood gases									
• PaO ₂ - mmHg	181	74.5 ± 14.1	57	74.8 ± 14.2	91	72.9 ± 13.7	74.7 ± 13.2	0.42	0.37
• PaCO ₂ - mmHg	182	39.6 ± 4.8	57	40.0 ± 4.8	90	39.8 ± 4.9	40.0 ± 6.0	0.86	0.86
• pH	182	7.42 ± 0.03	57	7.41 ± 0.03	92	7.41 ± 0.03	7.41 ± 0.04	0.78	0.70

<i>Respiratory function</i>										
• Pre-bronchodilator FEV ₁ (% predicted)	233	59.6 ± 24.3	72	63.0 ± 27.7	144	56.6 ± 21.7	56.1 ± 21.7	0.10	0.84	
• Post-bronchodilator FEV ₁ (% predicted)	171	63.7 ± 25.8	53	71.3 ± 30.4	70	56.6 ± 22.8	56.5 ± 21.5	0.005	0.98	
• With fixed airflow obstruction – no. (%) §	176	143 (81.3)	55	34 (61.8)	70	65 (92.9)	65 (92.9)	< 0.0001	1.00	
• Pre-bronchodilator FVC (% predicted)	233	86.6 ± 21.3	73	86.1 ± 21.9	142	85.0 ± 20.4	84.0 ± 20.4	0.73	0.68	
• Post-bronchodilator FVC (% predicted)	169	91.5 ± 21.1	51	91.9 ± 21.5	69	91.0 ± 20.4	88.8 ± 21.1	0.83	0.54	
• Pre-bronchodilator FEV ₁ / FVC (% predicted)	235	53.5 ± 15.3	74	57.0 ± 16.7	144	51.4 ± 14.0	51.7 ± 14.2	0.01	0.89	
• Post-bronchodilator FEV ₁ / FVC (% predicted)	176	53.9 ± 16.2	55	60.6 ± 17.3	70	47.2 ± 14.4	48.4 ± 14.7	< 0.0001	0.62	
• FRC - %	124	137.8 ± 38.3	49	131.5 ± 31.7	37	140.9 ± 43.4	139.9 ± 41.5	0.27	0.92	
• RV - %	135	156.4 ± 53.0	53	150.7 ± 43.1	41	165.9 ± 62.9	154.7 ± 56.6	0.19	0.41	
• TLC - %	137	113.8 ± 19.2	55	112.0 ± 16.9	42	114.1 ± 21.7	112.4 ± 19.5	0.60	0.70	
• DLCO - %	115	59.0 ± 21.7	46	63.9 ± 21.6	34	49.5 ± 18.8	49.9 ± 18.8	0.003	0.94	
<i>Symptoms</i>										
• With cough – no. (%)	241	97 (40.2)	78	27 (34.6)	163	70 (42.9)	56 (34.4)	0.22	0.12	
• With wheezing – no. (%)	241	32 (13.3)	78	14 (17.9)	163	18 (11.0)	16 (9.8)	0.14	0.72	
• With emphysema no. (%)	241	92 (38.2)	78	28 (35.9)	163	64 (39.3)	64 (39.3)	0.62	1.00	
• With exacerbations in the previous 12 months - no. (%)	241	128 (53.1)	78	39 (50.0)	163	89 (54.6)	62 (38.0)	0.51	0.003	
• Exacerbations in the previous 12 months - no.	241	1.44 ± 0.16	78	1.26 ± 0.22	163	1.52 ± 0.21	1.46 ± 0.23	0.12	0.72	
• Unscheduled medical visits in the previous 12 months - no.	241	1.19 ± 0.17	78	0.96 ± 0.23	163	1.29 ± 0.22	1.08 ± 0.16	0.03	0.09	
• Hospitalizations for COPD in the previous 12 months – no.	241	0.32 ± 0.05	78	0.41 ± 0.11	163	0.27 ± 0.05	0.19 ± 0.06	0.10	0.17	
<i>Comorbidities</i>										
• Cardiovascular – no. (%)	241	103 (42.7)	78	32 (41.0)	163	71 (43.6)	-	0.71	-	
• Hypertension – no. (%)	241	67 (27.8)	78	21 (26.9)	163	46 (28.2)	-	0.84	-	
• Dyslipidemia – no. (%)	241	58 (24.1)	78	14 (17.9)	163	44 (27.0)	-	0.13	-	
• Diabetes – no. (%)	241	28 (11.6)	78	7 (9.0)	163	21 (12.9)	-	0.38	-	
• Sinusitis – no. (%)	241	27 (11.2)	78	7 (9.0)	163	20 (12.3)	-	0.45	-	
• Allergic rhinitis – no. (%)	241	38 (15.8)	78	12 (15.4)	163	26 (15.9)	-	0.92	-	
• Obstructive sleep apnea – no. (%)	241	22 (9.1)	78	4 (5.1)	163	18 (11.0)	-	0.14	-	
• Gastro Esophageal Reflux – no. (%)	241	70 (29.0)	78	21 (26.9)	163	49 (30.1)	-	0.62	-	
• Bronchial dilatation – no. (%)	241	11 (4.6)	78	3 (3.8)	163	8 (4.9)	-	0.72	-	

<i>Treatments</i>										
• On SABA – no. (%)	231	144 (62.3)	70	33 (47.1)	158	110 (69.6)	115 (72.8)	0.002	0.54	
• On LABA alone – no. (%)	233	26 (11.2)	70	5 (7.1)	160	21 (13.1)	18 (11.2)	0.19	0.61	
• On LAMA alone – no. (%)	233	16 (6.9)	70	4 (5.7)	160	12 (7.5)	12 (7.5)	0.63	1.00	
• On ICS alone – no. (%)	233	4 (1.7)	70	1 (1.4)	160	3 (1.9)	3 (1.9)	0.82	1.00	
Daily dose of ICS – µg of equivalents beclomethasone	119	1304 ± 669	28	1499 ± 811	77	1253 ± 633	1242 ± 626	0.11	0.92	
• On OCS – no. (%)	232	8 (3.4)	69	7 (10.1)	160	1 (0.6)	2 (1.2)	0.0003	0.57	
Daily dose of prednisone (mg)	8	20.4 ± 10.9	7	22.6 ± 9.6	1	5	5	0.14	-	
• On LABA + LAMA – no. (%)	233	22 (9.4)	70	6 (8.6)	160	16 (10.0)	21 (13.1)	0.74	0.39	
• On LABA + LAMA + ICS – no. (%)	233	78 (33.5)	70	22 (31.4)	160	55 (34.4)	63 (39.4)	0.67	0.36	
• On anti-histamine – no. (%)	241	20 (8.3)	78	8 (10.3)	162	12 (7.4)	8 (4.9)	0.46	0.36	
• On theophylline – no. (%)	232	1 (0.4)	69	0 (0.0)	160	1 (0.6)	1 (0.6)	0.52	1.00	
• On gastro esophageal reflux inhibitors – no. (%)	241	56 (23.2)	78	17 (21.8)	163	39 (23.9)	29 (17.8)	0.72	0.18	
• On anti-hypertensive drugs – no. (%)	232	80 (34.5)	70	24 (34.3)	160	55 (34.4)	54 (33.7)	0.99	0.91	
• On statins – no. (%)	231	56 (24.2)	69	11 (15.9)	160	44 (27.5)	49 (30.6)	0.07	0.54	
• On fibrates – no. (%)	229	5 (2.2)	69	1 (1.4)	157	4 (2.5)	3 (1.9)	0.61	0.71	
• On anti-platelet aggregation – no. (%)	231	63 (27.3)	69	15 (21.7)	159	47 (29.6)	42 (26.4)	0.23	0.54	
• Other – no. (%)	227	102 (44.9)	68	20 (29.4)	157	81 (51.6)	75 (47.8)	0.003	0.50	
• On oxygen therapy – no. (%)	238	37 (15.5)	75	11 (14.7)	161	26 (16.1)	16 (9.9)	0.78	0.10	
Adherence to treatments – no. (%)	224	203 (90.6)	64	55 (85.9)	155	144 (92.9)	143 (92.3)	0.11	0.83	

Abbreviations: CRP, C reactive protein; FEV₁, Forced Expiratory Volume in 1 second; FVC, Forced Vital Capacity; FRC, Functional Residual Capacity; RV, Residual Volume; TLC, Total Lung Capacity; DLCO, transfer factor of the lung for carbon monoxide; ICS, inhaled corticosteroids; SABA, short-acting β₂-agonists; LABA, long-acting β₂-agonists; OCS, oral corticosteroids; LAMA, long-lasting muscarinic antagonists.

* denotes the number of patients with each available variable

† $P \leq 0.05$, between visits 1 for n=78 and n=163 COPD patients, with one and two visits, respectively (Students' t test, or Mann-Whitney U-test, or Fisher exact test, 2-tailed, or Poisson test)

‡ $P \leq 0.05$, between visits 1 and 2 for n=163 COPD patients (Students' t test, or Mann-Whitney U-test, or Fisher exact test, 2-tailed, or Poisson test)

§ With post-bronchodilator FEV₁/FVC lower than of 70% of predicted

|| Diagnosis of emphysema was established by CT scan assessment and/or by the measurement of DLCO

Data are n (%), or means ± SD, or median (25-75 % CI), or means ± SEM, in case of number of exacerbations, of unscheduled hospital visits and hospitalizations for COPD in the previous 6 months

Bold denotes statistical significance

Table 4: characteristics of COPD patients of the MLCC cohort

Parameter	n*	All patients, n = 47
Male sex – no. (%)	47	29 (61.7)
Age (years)	47	72.5 ± 9.6
Caucasian origin – no. (%)	47	47 (100)
Other origin – no. (%)	47	0
<i>GOLD stages</i>		
▪ GOLD 1 – no. (%)	47	2 (4.3)
▪ GOLD 2 – no. (%)	47	16 (34.0)
▪ GOLD 3 – no. (%)	47	19 (40.4)
▪ GOLD 4 – no. (%)	47	10 (21.3)
<i>Smoking history</i>		
▪ Never smokers – no. (%)	47	0
▪ Former smoker – no. (%)	47	41 (87.2)
▪ Packs per year in former smokers – no.	47	56.9 ± 32.6
▪ Active smokers – no. (%)	47	6 (12.8)
▪ Packs per year in active smokers – no.	47	89.2 ± 52.1
Body Mass Index (kg per m ²)	31	25.2 ± 6.0
<i>Biology</i>		
▪ Blood leukocytes (no. per mm ³)	45	7800 (6500 - 9800)
▪ Blood eosinophils (no. per mm ³)	45	200 (100 - 300)
▪ With blood eosinophils ≥ 300 per mm ³ – no. (%)	45	14 (31.1)
▪ Blood neutrophils (no. per mm ³)	45	5300 (4300 - 6400)
▪ Blood lymphocytes (no. per mm ³)	45	1600 (1100 - 2000)
▪ Blood monocytes (no. per mm ³)	45	700 (600 - 800)
▪ Hemoglobin – g per deciliter	45	13.1 ± 1.6
▪ CRP – mg per Liter	34	1.4 ± 2.9
▪ With CRP ≥ 3 mg per Liter – no. (%)	34	2 (5.9)
<i>Blood gases</i>		
▪ PaO ₂ - mmHg	9	66.1 ± 5.4
▪ PaCO ₂ - mmHg	9	40.5 ± 4.6
<i>Respiratory function</i>		
▪ Pre-bronchodilator FEV ₁ (% predicted)	28	39.1 ± 10.5
▪ Post-bronchodilator FEV ₁ (% predicted)	29	41.9 ± 14.2
▪ Pre-bronchodilator FVC (% predicted)	28	87.1 ± 16.9
▪ Post-bronchodilator FVC (% predicted)	30	93.1 ± 18.0
▪ Pre-bronchodilator FEV ₁ / FVC (% predicted)	28	35.5 ± 9.0
▪ Post-bronchodilator FEV ₁ / FVC (% predicted)	29	35.5 ± 10.0
▪ With fixed airflow obstruction – no. (%) ^a	29	28 (96.6)
▪ DLCO (%)	25	41.4 ± 12.4
<i>Symptoms</i>		
▪ With cough – no (%)	46	38 (82.6)
▪ With wheezing – no (%)	47	7 (14.9)
With emphysema no. (%) ^b	25	24 (96.0)
Exacerbations in 12 months - no. (%) ^c	47	1.6 ± 0.7
Hospitalizations for COPD in 12 months - no. (%) ^c	17	0.7 ± 0.7

<i>Comorbidities</i>		
▪ Cardiovascular – no. (%)	47	22 (46.8)
▪ Diabetes – no. (%)	47	9 (19.1)
▪ Obstructive sleep apnea – no. (%)	47	1 (2.1)
<i>Treatments</i>		
On short-acting β 2-agonists – no. (%)	47	44 (93.6)
On inhaled steroids – no. (%)	47	43 (91.5)
▪ Daily dose of inhaled steroids - μ g of EQB	40	842 $\pm$ 366
On oral steroids – no. (%)	47	4 (8.5)
▪ Daily dose of prednisone (mg)	4	4.9 $\pm$ 2.3
On anti-cholinergics – no. (%)	43	42 (97.7)
On oxygen therapy – no. (%)	47	27 (57.4)
Adherence to treatment – no. (%)	12	8 (66.7)

Abbreviations: CRP, C reactive protein; FEV₁, Forced Expiratory Volume in 1 second; FVC, Forced Vital Capacity; DLCO, transfer factor of the lung for carbon monoxide; EQB, equivalents of beclomethasone

^a With post-bronchodilator FEV1 lower than of 70% of predicted

^b Diagnosis of emphysema was established by the measurement of DLCO

^c Estimated frequency of exacerbations and hospitalizations based on the events during the study period

* denotes the number of patients with each available variable

Data are n (%), or means $\pm$ SD, or median (25-75 IQR)

5

10

15

20

Abbreviations: CRP, C reactive protein; FEV₁, Forced Expiratory Volume in 1 second; FVC, Forced Vital Capacity; FRC, Functional Residual Capacity; RV, Residual Volume; TLC, Total Lung Capacity

* indicates the number of patients with each available variable

Data are expressed as numbers (%), as means $\pm$ SD or SEM, or as medians (25-75 IQR)

^a Students' t test, Fisher exact test 2-tailed and Mann-Whitney-U test

Bold denotes statistical significance

5 **Table 5: Significant correlations between differentially regulated proteins and the incidence of exacerbations and emphysema in patients with COPD.**

Biological Hallmarks	Target	Exacerbations (Y/N)		Emphysema (Y/N)	
		OR (2.5-97.5%)	FDR	OR (2.5-97.5%)	FDR
Cell fate, Remodeling and Repair	PRKCA *	0.57 (0.41-0.79)	0.016	-	-
	PRKACA *	-	-	0.62 (0.41-0.90)	0.0200
	SHC1	-	-	0.33 (0.17-0.58)	0.0026
	eIF-5A-1	-	-	0.43 (0.18-0.60)	0.0026
	TCTP	0.57 (0.36-0.87)	0.0476	0.35 (0.20-0.58)	0.0026
	PA2G4	-	-	0.40 (0.24-0.64)	0.0026
	14-3-3 protein β/α *	0.50 (0.30-0.79)	0.0281	0.41 (0.24-0.68)	0.0032
	14-3-3 protein ζ/δ *	0.45 (0.28-0.70)	0.0160	0.42 (0.26-0.68)	0.0031
	BAD *	-	-	0.45 (0.27-0.70)	0.0031
	RPS6KA3 *	0.61 (0.40-0.89)	0.048	0.46 (0.29-0.71)	0.0031
	AREG	-	-	0.46 (0.27-0.75)	0.0065
	METAP1 *	-	-	0.50 (0.32-0.76)	0.0044
	FGF16	-	-	0.52 (0.31-0.81)	0.0115
	RAC1 *	0.58 (0.40-0.82)	0.022	0.52 (0.35-0.76)	0.0033
	SBDS	0.66 (0.48-0.92)	0.048	0.53 (0.36-0.76)	0.0031

	hnRNP A2/B1	-	-	0.53 (0.36-0.79)	0.0060
	NSF1C	-	-	0.55 (0.39-0.77)	0.0031
	PKC-A *	-	-	0.55 (0.38-0.78)	0.0042
	SMAD2 *	0.67 (0.49-0.92)	0.048	0.58 (0.40-0.81)	0.0059
	eIF-4H	0.63 (0.45-0.86)	0.028	0.59 (0.41-0.82)	0.0061
	DRG-1	-	-	0.51 (0.35-0.72)	0.0026
	DLRB1	-	-	0.61 (0.39-0.91)	0.025
	PLPP	-	-	0.36 (0.20-0.62)	0.0026
	ARGI1	0.62 (0.43-0.88)	0.047	-	-
	IMB1 *	-	-	0.60 (0.43-0.80)	0.003
	ARGI1 *	0.62 (0.43-0.88)	0.047	-	-
	41	-	-	0.62 (0.44-0.86)	0.010
	GRB2 adapter protein	0.72 (0.55-0.93)	0.017	0.64 (0.49-0.85)	0.004
	Tropomyosin 4	0.78 (0.64-0.96)	0.024	0.71 (0.57-0.89)	0.004
	Sphingosine kinase 1	0.66 (0.47-0.91)	0.048	0.60 (0.41-0.85)	0.008
	Haemoglobin	-	-	0.72 (0.54-0.93)	0.027
	CSK	-	-	0.64 (0.47-0.86)	0.006
	FYN *	-	-	0.74 (0.56-0.97)	0.042
	ERK-1 *	0.50 (0.29-0.84)	0.0476	0.41 (0.22-0.72)	0.0062
	MK01/ERK-2	-	-	0.45 (0.26-0.74)	0.0061
	PKC-B-II	0.59 (0.43-0.80)	0.0158	0.55 (0.40-0.76)	0.0026

	Prostatic binding protein	-	-	0.41 (0.24-0.68)	0.0032
	Caspase-3	0.66 (0.47-0.91)	0.048	0.52 (0.35-0.74)	0.0030
	PDPK1 *	-	-	0.61 (0.43-0.87)	0.011
	LYN *	0.65 (0.47-0.87)	0.029	0.65 (0.47-0.89)	0.013
	LYNB *	0.65 (0.48-0.87)	0.028	0.68 (0.50-0.91)	0.016
	CK2-A1:B	-	-	0.56 (0.40-0.80)	0.0164
	NDP kinase B *	-	-	0.52 (0.36-0.74)	0.0026
	FER *	0.71 (0.55-0.91)	0.016	0.67 (0.51-0.87)	0.004
	RAN *	-	-	0.60 (0.42-0.82)	0.0063
Metabolism and Mitochondria	RAC1	0.58 (0.40-0.82)	0.022	0.52 (0.35-0.76)	0.0033
	□-Synuclein *	-	-	0.53 (0.36-0.75)	0.0031
	Cyclophilin F	0.72 (0.57-0.91)	0.016	0.63 (0.50-0.84)	0.003
	Cyclophilin A *	0.42 (0.24-0.73)	0.0222	0.31 (0.16-0.57)	0.0026
	SUMO3	-	-	0.38 (0.21-0.63)	0.0026
	UFM1	-	-	0.48 (0.31-0.70)	0.0026
	SNAA	-	-	0.47 (0.31-0.69)	0.0026
	VAV *	0.70 (0.53-0.90)	0.040	0.66 (0.49-0.86)	0.006
	14-3-3 protein β/α *	0.50 (0.30-0.79)	0.0281	0.41 (0.24-0.68)	0.0032
	14-3-3 protein ζ/δ *	0.45 (0.28-0.70)	0.0160	0.42 (0.26-0.68)	0.0031
	BAD	-	-	0.45 (0.27-0.70)	0.0031
	Triosephosphate isomerase	-	-	0.37 (0.21-0.61)	0.0026
	RPS6KA3	0.61 (0.40-0.89)	0.048	0.46 (0.29-0.71)	0.0031
	METAP1	-	-	0.50 (0.32-0.76)	0.0044

Ubiquitin+1 *	-	-	0.37 (0.10-0.66)	0.0045
UBC9	-	-	0.38 (0.21-0.67)	0.0026
UBE2N *	-	-	0.46 (0.31-0.68)	0.0026
Carbonic anhydrase XIII	0.80 (0.65-0.99)	0.038	0.74 (0.59-0.94)	0.011
SRCN1	0.64 (0.49-0.83)	0.016	0.75 (0.57-0.97)	0.035
M2-PK	0.55 (0.39-0.77)	0.016	0.72 (0.53-0.98)	0.045
PPAC	-	-	0.64 (0.47-0.85)	0.006
Transketolase	-	-	0.38 (0.22-0.63)	0.0026
Aflatoxin B1 aldehyde reductase	-	-	0.61 (0.43-0.84)	0.007
NCC27	-	-	0.46 (0.25-0.80)	0.0129
PPID	0.60 (0.43-0.81)	0.016	0.70 (0.51-0.95)	0.033
Myokinase	-	-	0.64 (0.48-0.85)	0.006
NACA	-	-	0.53 (0.35-0.79)	0.0061
SGTA	-	-	0.64 (0.46-0.88)	0.013
H2A3	0.87 (0.69-1.09)	0.2495	0.81 (0.63-1.04)	0.014
BTK *	0.72 (0.57-0.92)	0.016	0.63 (0.49-0.82)	0.003
RPS6KA3	0.61 (0.40-0.89)	0.048	0.46 (0.29-0.71)	0.0031
DUS3	-	-	0.50 (0.30-0.78)	0.0072
RAC1 *	0.58 (0.40-0.82)	0.022	0.52 (0.35-0.76)	0.0033
CD40 ligand	-	-	0.54 (0.35-0.81)	0.0074
CPNE1 *	0.56 (0.35-0.88)	0.048	0.51 (0.30-0.83)	0.0142
IMB1	-	-	0.60 (0.43-0.80)	0.003
PTP-1C	-	-	0.63 (0.42-0.91)	0.023
SRCN1	0.64 (0.49-0.83)	0.016	0.75 (0.57-0.97)	0.035
Annexin I	-	-	0.59 (0.39-0.88)	0.0164
BARK1	-	-	0.62 (0.41-0.92)	0.026
GAPDH	0.64 (0.47-0.84)	0.022	0.59 (0.44-0.78)	0.003
SP-D	-	-	0.76 (0.59-0.97)	0.034
STAT3	0.47 (0.26-0.83)	0.0476	0.38 (0.20-0.71)	0.0026

	6-Phosphoglucose dehydrogenase	0.68 (0.53-0.85)	0.016	-	-
	ARGI1	0.62 (0.43-0.88)	0.047	-	-
	Midkine	-	-	1.5 (1.1-2.0)	0.014
	Lactadherin (MFGM)	-	-	1.9 (1.3-2.9)	0.006
Tissue injury	MMP-12	-	-	1.51 (1.07-2.15)	0.027
	Renin	-	-	1.88 (1.26-2.87)	0.006

Abbreviations: OR, odd-ratio; FDR, false discovery rate. For protein abbreviations, see legend of Table 1 of the main manuscript. * These proteins have multiples biological functions and, therefore, they belong to more than one biological process

Table 6: Main clinical characteristics of COPD patients switching of Cluster between the first visit (inclusion) and the follow up visit

Parameter	Values at Visit 1	Values at Visit 2	P value *
COPD patients switching from Cluster 1 at Visit 1 to Cluster 2 at Visit 2 (n = 29)			
With emphysema - no. (%) †	16 (55.2)	13 (44.8)	0.44
DLCO - %	49.4 ± 13.3	54.3 ± 28.5	0.60
With exacerbations in the previous 12 months - no. (%)	20 (69.0)	11 (37.9)	0.02
With unscheduled medical visits in the previous 12 months - no. (%)	16 (55.2)	8 (27.6)	0.04
With hospitalizations for COPD in the previous 12 months - no. (%)	8 (27.6)	1 (3.4)	0.02
On SABA - no. (%)	19 (65.5)	18 (69.2)	1.00
On LABA - no. (%)	22 (75.9)	20 (70)	0.77
On LAMA - no. (%)	19 (65.5)	20 (70)	1.00
On ICS - no. (%)	13 (44.8)	15 (51.7)	0.79
On anti-hypertensive drugs - no. (%)	12 (41.4)	10 (38.5)	0.79
On statins - no. (%)	11 (37.9)	10 (38.5)	1.00
Adherence to treatment - no. (%)	26 (89.7)	26 (89.7)	1.00
COPD patients switching from Cluster 2 at visit 1 to Cluster 1 at visit 2 (n = 26)			
With emphysema - no. (%) †	5 (19.2)	11 (42.3)	0.08
DLCO - %	53.6 ± 18.0	53.4 ± 10.9	0.99
With exacerbations in the previous 12 months - no. (%)	15 (57.7)	10 (38.5)	0.17
With unscheduled medical visits in the previous 12 months - no. (%)	10 (38.5)	8 (30.8)	0.56
With hospitalizations for COPD in the previous 12 months - no. (%)	1 (4)	6 (23.1)	0.05
On SABA - no. (%)	20 (69)	19 (73.1)	1.00
On LABA - no. (%)	18 (62.1)	21 (80.8)	0.52
On LAMA - no. (%)	12 (46.2)	20 (76.9)	0.04
On ICS - no. (%)	13 (50)	15 (57.7)	0.74
On anti-hypertensive drugs - no. (%)	7 (24.1)	8 (30.8)	1.00
On statins - no. (%)	2 (6.9)	3 (11.5)	1.00
Adherence to treatments - no. (%)	23 (92.0)	24 (100.0)	1.00

Abbreviation: DLCO, transfer factor of the lung for carbon monoxide

Data are expressed as numbers (%) and as means ± SD

* Students' t test or Fisher exact test 2-tailed

† Diagnosis of emphysema was established by CT scan and/or by the measurement of DLCO

Bold denotes statistical significance

REFERENCES:

Throughout this application, various references describe the state of the art to which this invention pertains. The disclosures of these references are hereby incorporated by reference into the present disclosure.

1. Vogelmeier CF, Criner GJ, Martinez FJ, Anzueto A, Barnes PJ, Bourbeau J, Celli BR,
5 Chen R, Decramer M, Fabbri LM, et al. Global Strategy for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease 2017 Report: GOLD Executive Summary. *Eur Respir J*. 2017;49: 1700214.
2. Aggarwal T, Wadhwa R, Thapliyal N, Sharma K, Rani V, Maurya PK. Oxidative, inflammatory, genetic, and epigenetic biomarkers associated with chronic obstructive
10 pulmonary disorder. *J Cell Physiol*. 2019;234:2067–2082.
3. Brandsma C-A, de Vries M, Costa R, Woldhuis RR, Königshoff M, Timens W. Lung ageing and COPD: is there a role for ageing in abnormal tissue repair? *Eur Respir Rev Off J Eur Respir Soc*. 2017;26(146).
4. Barnes PJ. Senescence in COPD and Its Comorbidities. *Annu Rev Physiol*.
15 2017;79:517–39.
5. Segal LN, Martinez FJ. Chronic obstructive pulmonary disease subpopulations and phenotyping. *J Allergy Clin Immunol*. 2018;141:1961–1971.
6. Brown JP, Martinez CH. Chronic obstructive pulmonary disease comorbidities. *Curr Opin Pulm Med*. 2016;22:113–118.
- 20 7. Agustí A, Celli B, Faner R. What does endotyping mean for treatment in chronic obstructive pulmonary disease? *Lancet Lond Engl*. 2017;390:980–987.
8. Sidhaye VK, Nishida K, Martinez FJ. Precision medicine in COPD: where are we and where do we need to go? *Eur Respir Rev Off J Eur Respir Soc*. 2018;27(149).
9. Jensen AB, Moseley PL, Oprea TI, Ellesøe SG, Eriksson R, Schmock H, Jensen PB,
25 Jensen LJ, Brunak S. Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. *Nat Commun*. 2014;5:4022.
10. Zhou A, Zhou Z, Zhao Y, Chen P. The recent advances of phenotypes in acute exacerbations of COPD. *Int J Chron Obstruct Pulmon Dis*. 2017;12:1009–1018.
11. Subramanian DR, Gupta S, Burggraf D, Vom Silberberg SJ, Heimbeck I, Heiss-
30 Neumann MS, Haeussinger K, Newby C, Hargadon B, Raj V et al. Emphysema- and airway-dominant COPD phenotypes defined by standardised quantitative computed tomography. *Eur Respir J*. 2016;48:92–103.

12. Keene JD, Jacobson S, Kechris K, Kinney GL, Foreman MG, Doerschuk CM, Make BJ, Curtis JL, Rennard SI, Barr RG et al. Biomarkers Predictive of Exacerbations in the SPIROMICS and COPD Gene Cohorts. *Am J Respir Crit Care Med*. 2017;195:473–481.
13. Gold L, Ayers D, Bertino J, Bock C, Bock A, Brody EN, Carter J, Dalby AB, Eaton BE, Fitzwater T, et al. Aptamer-based multiplexed proteomic technology for biomarker discovery. *PloS One*. 2010;5:e15004.
14. DeBoer EM, Wagner BD, Popler J, Harris JK, Zemanick ET, Accurso FJ, Sagel SD, Deterding RR. Novel Application of Aptamer Proteomic Analysis in Cystic Fibrosis Bronchoalveolar Lavage Fluid. *Proteomics Clin Appl*. 2019;13:e1800085.
- 10 15. Deterding RR, Wagner BD, Harris JK, DeBoer EM. Pulmonary Aptamer Signatures in Children's Interstitial and Diffuse Lung Disease. *Am J Respir Crit Care Med*. 2019;200:1496–1504.
16. Norman KC, O'Dwyer DN, Salisbury ML, DiLillo KM, Lama VN, Xia M, Gurczynski SJ, White ES, Flaherty KR, Martinez FJ. Identification of a unique temporal signature in blood and BAL associated with IPF progression. *Sci Rep*. 2020;10:12049.
- 15 17. Bourdin A, Suehs CM, Marin G, Vachier I, Matzner-Lober E, Chanez P, Molinari N; COBRA Consortium. Asthma, COPD, and overlap in a national cohort: ACO on a gradient. *J Allergy Clin Immunol*. 2018;141:1516–1518.
18. Pretolani M, Soussan D, Poirier I, Thabut G, Aubier M, COBRA Study Group, Clinical and biological characteristics of the French COBRA cohort of adult subjects with asthma. *Eur Respir J*. 2017;50(2).
- 20 19. Hutchinson AF, Ghimire AK, Thompson MA, Black JF, Brand CA, Lowe AJ, Smallwood DM, Vlahos R, Bozinovski S, Brown GV et al. A community-based, time-matched, case-control study of respiratory viruses and exacerbations of COPD. *Respir Med*. 2007;101:2472–2481.
- 25 20. Rohloff JC, Gelinas AD, Jarvis TC, Ochsner UA, Schneider DJ, Gold L, Janjic N. Nucleic Acid Ligands With Protein-like Side Chains: Modified Aptamers and Their Use as Diagnostic and Therapeutic Agents. *Mol Ther Nucleic Acids*. 2014;3:e201.
21. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43:e47.
- 30 22. Kelsen SG. The Unfolded Protein Response in Chronic Obstructive Pulmonary Disease. *Ann Am Thorac Soc*. 2016; Suppl 2:S138-S145.

23. Xiong H, Yan J, Cai S, He Q, Peng D, Liu Z, Liu Y. Cancer protein biomarker discovery based on nucleic acid aptamers. *Int J Biol Macromol*. 2019;132:190–202.
24. Pastor MD, Nogal A, Molina-Pinelo S, Meléndez R, Salinas A, González De la Peña M, Martín-Juan J, Corral J, García-Carbonero R, Carnero A et al. Identification of proteomic signatures associated with lung cancer and COPD. *J Proteomics*. 2013;89:227–237.
25. Hu Y, Li L, Shen L, Gao H, Yu F, Yin W, Liu W. FGF-16 protects against adverse cardiac remodeling in the infarct diabetic heart. *Am J Transl Res*. 2017;9:1630–1640.
26. Yu W, Huang X, Tian X, Zhang H, He L, Wang Y, Nie Y, Hu S, Lin Z, Zhou B et al. GATA4 regulates Fgf16 to promote heart repair after injury. *Dev Camb Engl*. 2016;143:936–949.
27. Davies DE, Polosa R, Puddicombe SM, Richter A, Holgate ST. The epidermal growth factor receptor and its ligand family: their potential role in repair and remodelling in asthma. *Allergy*. 1999;54:771–783.
28. Zheng Y, Zhang C, Croucher DR, Soliman MA, St-Denis N, Pasculescu A, Taylor L, Tate SA, Hardy WR, Colwill K et al. Temporal regulation of EGF signalling networks by the scaffold protein Shc1. *Nature*. 2013;499:166–171.
29. Levine S, Kaiser L, Leferovich J, Tikunov B. Cellular adaptations in the diaphragm in chronic obstructive pulmonary disease. *N Engl J Med*. 1997;337:1799–1806.
30. Gunning P, O'Neill G, Hardeman E. Tropomyosin-based regulation of the actin cytoskeleton in time and space. *Physiol Rev*. 2008;88:1–35.
31. Skeletal muscle dysfunction in chronic obstructive pulmonary disease. A statement of the American Thoracic Society and European Respiratory Society. *Am J Respir Crit Care Med*. 1999;159(4 Pt 2):S1-S40.
32. Barreiro E, Gea J, Matar G, Hussain SNA. Expression and carbonylation of creatine kinase in the quadriceps femoris muscles of patients with chronic obstructive pulmonary disease. *Am J Respir Cell Mol Biol*. 2005;33:636–642.
33. Bittner ZA, Liu X, Mateo Tortola M, Tapia-Abellán A, Shankar S, Andreeva L, Mangan M, Spalinger M, Kalbacher H, Düwell P et al. BTK operates a phospho-tyrosine switch to regulate NLRP3 inflammasome activity. *J Exp Med*. 2021;218:e20201656.
34. Florence JM, Krupa A, Booshehri LM, Gajewski AL, Kurdowska AK. Disrupting the Btk Pathway Suppresses COPD-Like Lung Alterations in Atherosclerosis Prone ApoE^{-/-} Mice Following Regular Exposure to Cigarette Smoke. *Int J Mol Sci*. 2018;19(2).

35. Lacy M, Bürger C, Shami A, Ahmadsei M, Winkels H, Nitz K, van Tiel CM, Seijkens TTP, Kusters PJH, Karshovka E et al. Cell-specific and divergent roles of the CD40L-CD40 axis in atherosclerotic vascular disease. *Nat Commun.* 2021;12:3754.
36. Nigro P, Pompilio G, Capogrossi MC. Cyclophilin A: a key player for human disease. *Cell Death Dis.* 2013;4:e888.
37. Zhang M, Tang J, Yin J, Wang X, Feng X, Yang X, Shan H, Zhang Q, Zhang J, Li Y. The clinical implication of serum cyclophilin A in patients with chronic obstructive pulmonary disease. *Int J Chron Obstruct Pulmon Dis.* 2018;13:357–363.
38. Şalaru DL, Arsenescu-Georgescu C, Chatzikyrkou C, Karagiannis J, Fischer A, Mertens PR. Midkine, a heparin-binding growth factor, and its roles in atherogenesis and inflammatory kidney diseases. *Nephrol Dial Transplant Off Publ Eur Dial Transpl Assoc - Eur Ren Assoc.* 2016;31:1781–1787.
39. Bădilă E, Daraban AM, Țintea E, Bartoș D, Alexandru N, Georgescu A. Midkine proteins in cardio-vascular disease. Where do we come from and where are we heading to? *Eur J Pharmacol.* 2015;762:464–471.
40. Lee HA, Lim J, Joo HJ, Lee Y-S, Jung YK, Kim JH, Byun KS, Seo YS. Serum milk fat globule-EGF factor 8 protein as a potential biomarker for metabolic syndrome. *Clin Mol Hepatol.* 2021;27:463–473.
41. Hao W, Li M, Zhang Y, Zhang C, Xue Y. Expressions of MMP-12, TIMP-4, and Neutrophil Elastase in PBMCs and Exhaled Breath Condensate in Patients with COPD and Their Relationships with Disease Severity and Acute Exacerbations. *J Immunol Res.* 2019;7142438.
42. Kuba K, Imai Y, Penninger JM. Angiotensin-converting enzyme 2 in lung diseases. *Curr Opin Pharmacol.* 2006;6:271–276.
43. Yang L, Zhou F, Zheng D, Wang D, Li X, Zhao C, Huang X. FGF/FGFR signaling: From lung development to respiratory diseases. *Cytokine Growth Factor Rev.* 2021;S1359-S6101(21)00069-1.
44. Cochran DL, Oh T-J, Mills MP, Clem DS, McClain PK, Schallhorn RA, McGuire MK, Scheyer ET, Giannobile WV, Reddy MS et al. A Randomized Clinical Trial Evaluating rh-FGF-2/ β -TCP in Periodontal Defects. *J Dent Res.* 2016;95:523–530.
45. Shah SS, Mithoefer K. Current Applications of Growth Factors for Knee Cartilage Repair and Osteoarthritis Treatment. *Curr Rev Musculoskelet Med.* 2020;13:641–650.

CLAIMS:

1. A method of predicting the clinical outcome of a patient suffering from chronic obstructive pulmonary disease (COPD) comprising the steps consisting of determining the level of a least one protein selected from the group consisting of BTK, Cyclophilin F, Tropomyosin 4, FGF-16, AREG, SHC1, 14-3-3 protein β/α , eIF-4H, CD40 ligand, H2A3, Midkine, MFGM, MMP-12, Renin and Carbonic anhydrase XIII wherein said level indicates the clinical outcome.
2. The method of claim 1 wherein the level of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15 proteins is determined in the blood sample.
3. The method of claim wherein high levels of: AREG, FGF-16, SHC1, 14-3-3 protein β/α , eIF-4H, Tropomyosin 4, Cyclophilin F, Carbonic anhydrase XIII, H2A3, CD40 ligand, BTK and low levels of: midkine, Lactadherin, MMP-12, renin indicate a low risk of worsening and/or high levels of: Midkine, Lactadherin, MMP-12, renin and low levels of: AREG, FGF-16, SHC1, 14-3-3 protein β/α , eIF-4H, Tropomyosin 4, Cyclophilin F, Carbonic anhydrase XIII, H2A3, CD40 ligand, BTK indicate a high risk of worsening.
4. The method of claim 1 that further comprises comparing the level of protein with a predetermined reference value wherein detecting a difference between the level of the protein and the predetermined reference value indicates the clinical outcome (e.g. the risk of worsening).
5. The method of claim 1 that further comprises a step consisting of calculating a score, representing an estimation of the clinical outcome and/or risk of worsening.
6. The method of claim 1 that further comprises a) determining the level of one or more protein(s) in the sample obtained from the subject; b) implementing an algorithm on data comprising the level of the protein(s) so as to obtain an algorithm output; c) determining the clinical outcome and/or the risk of worsening.

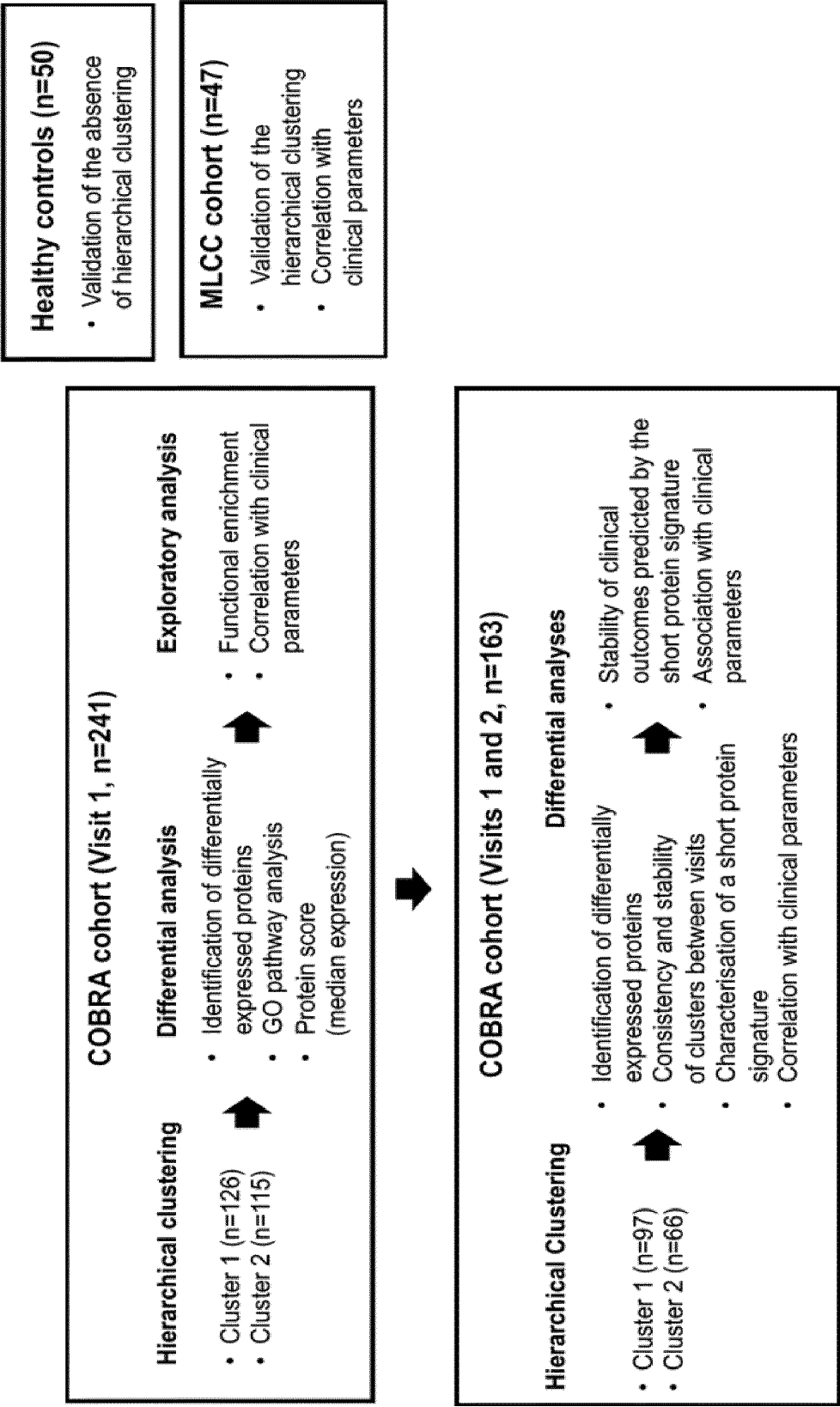


Figure 1

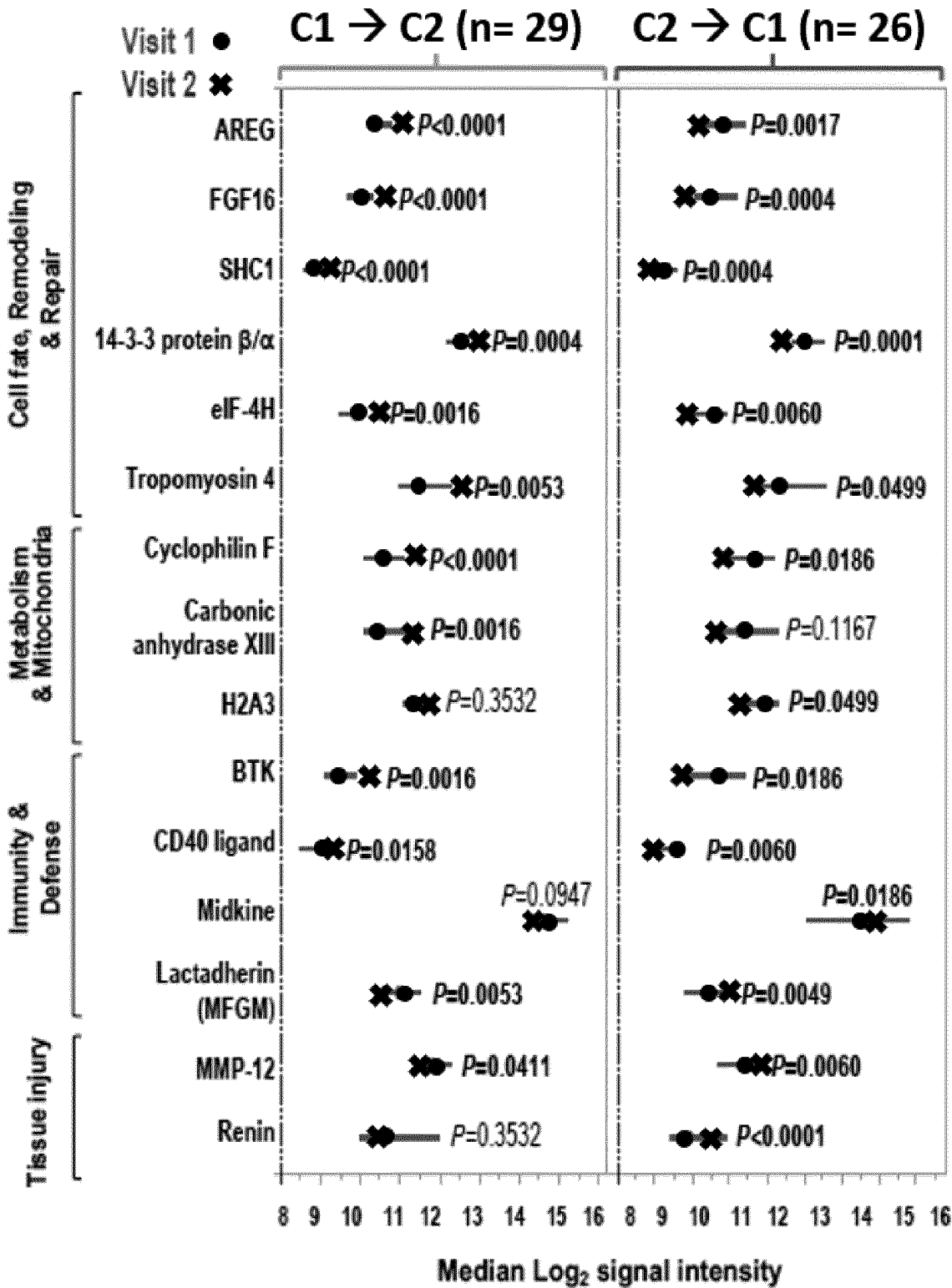


Figure 2

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2023/062030**A. CLASSIFICATION OF SUBJECT MATTER**
INV. G01N33/68
ADD.

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

B. FIELDS SEARCHEDMinimum documentation searched (classification system followed by classification symbols)
G01N

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

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

EPO-Internal**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	ABD-ELAZIZ KHALID ET AL: "Revisiting matrix metalloproteinase 12: its role in pathophysiology of asthma and related pulmonary diseases", CURRENT OPINION IN PULMONARY MEDICINE, vol. 27, no. 1, 16 October 2020 (2020-10-16), pages 54-60, XP055966127, US ISSN: 1070-5287, DOI: 10.1097/MCP.0000000000000743 the whole document In particular: Title; Abstract; "Key points"; Page 58, section about COPD. ----- -/--	1-6

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

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search

15 August 2023

Date of mailing of the international search report

16/10/2023Name and mailing address of the ISA/
European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

C.F. Angioni

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2023/062030

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CHEN RUIQI ET AL: "Prediction of prognosis in chronic obstructive pulmonary disease patients with respiratory failure: A comparison of three nutritional assessment methods", EUROPEAN JOURNAL OF INTERNAL MEDICINE, ELSEVIER, AMSTERDAM, NL, vol. 57, 12 June 2018 (2018-06-12), pages 70-75, XP085529933, ISSN: 0953-6205, DOI: 10.1016/J.EJIM.2018.06.006 the whole document In particular: Title; Abstract; Materials and methods; Discussion. -----	1-6
A	BRAJER B ET AL: "CONCENTRATION OF MATRIX METALLOPROTEINASE-9 IN SERUM OF PATIENTS WITH CHRONIC OBSTRUCTIVE PULMONARY DISEASE AND A DEGREE OF AIRWAY OBSTRUCTION AND DISEASE PROGRESSION", JOURNAL OF PHYSIOLOGY AND PHARMACOLOGY, vol. 59, no. 6 1 January 2008 (2008-01-01), pages 145-152, XP055966520, Retrieved from the Internet: URL:https://www.jpp.krakow.pl/journal/archive/12_08_s6/pdf/145_12_08_s6_article.pdf the whole document In particular: Title; Abstract; Materials and methods; Figure 1; Discussion last paragraph (conclusion). -----	1-6
X	FLORENCE JON ET AL: "Disrupting the Btk Pathway Suppresses COPD-Like Lung Alterations in Atherosclerosis Prone ApoE^{-/-} Mice Following Regular Exposure to Cigarette Smoke", INTERNATIONAL JOURNAL OF MOLECULAR SCIENCES, vol. 19, no. 2 24 January 2018 (2018-01-24), page 343, XP055838123, DOI: 10.3390/ijms19020343 Retrieved from the Internet: URL:https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5855565/pdf/ijms-19-00343.pdf the whole document In particular: Title; Abstract; Figures 1-9; Methods. -----	1-6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/EP2023/062030

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
 - a. ☒ forming part of the international application as filed.
 - b. ☐ furnished subsequent to the international filing date for the purposes of international search (Rule 13ter.1(a)).

☐ accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.
2. ☐ With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.
3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2023/062030

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

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

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

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

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims;; it is covered by claims Nos.:
1-6 (partially)

Remark on Protest

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

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-6(partially)

A method of predicting the clinical outcome of a patient suffering from chronic obstructive pulmonary disease (COPD) comprising the steps consisting of determining the level of BTK wherein said level indicates the clinical outcome.

2-16. claims: 1-6(partially)

A method of predicting the clinical outcome of a patient suffering from chronic obstructive pulmonary disease (COPD) comprising the steps consisting of determining the level of at least one protein selected from the group consisting of Cyclophilin F, Tropomyosin 4, FGF-16, AREG, SHC1, 14-3-3 protein ?/?, eIF-4H, CD40 ligand, H2A3, Midkine, MFGM, MMP-12, Renin and Carbonic anhydrase XIII wherein said level indicates the clinical outcome.
