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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 May 2014 (15.05.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/072086 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01) *G01N 33/574* (2006.01)

(21) International Application Number:
PCT/EP2013/060809

(22) International Filing Date:
24 May 2013 (24.05.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
12192103.5 9 November 2012 (09.11.2012) EP

(71) Applicant: **PHILIP MORRIS PRODUCTS S.A.** [CH/CH]; Quai Jeanrenaud 3, CH-2000 Neuchatel (CH).

(72) Inventors: **ANSARI, Sam**; Verger Clottu 3, CH-2068 Hauterive (CH). **GEBEL, Stephan**; Waldweg 4, 51519 Odenthal (DE). **GONZALEZ SUAREZ, Ignacio**; Grand Rue 7, CH-2035 Corcelles (CH). **HOENG, Julia**; Grand Rue 35, CH-2035 Corcelles (CH). **MARTIN, Florian**; Chemin de l'Orée 1, CH-2034 Peseux (CH). **PEITSCH, Manuel**; Chemin Gabriel 5, CH-2034 Peseux (CH). **XIANG, Yang**; Rue du Rocher 24, CH-2000 Neuchâtel (CH).

(74) Agent: **VOSSIUS & PARTNER (NO. 31)**; Siebertstraße 4, 81675 München (DE).

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

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

Published:

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

WO 2014/072086 A1

(54) Title: BIOMARKERS FOR PROGNOSIS OF LUNG CANCER

(57) Abstract: The present invention relates to panels of biomarkers that can be used to provide a prognosis of an individual's lung cancer and to predict and monitor the outcome of an individual's lung cancer which is or will be under treatment. The invention further provides kits for carrying out said methods.

Biomarkers for prognosis of lung cancer

The present invention relates to biomarkers that can be used to provide a prognosis of an individual's lung cancer and to predict and monitor the outcome of an individual's lung cancer which is or will be under treatment. The invention further provides kits for carrying out said methods.

Lung cancer is one of the most common cancers in the world and is a leading cause of cancer death in men and women in the developed world.

Diagnosis of lung cancer is still mainly based on chest radiograph and computed tomography (CT scan). Bronchoscopy or CT-guided biopsy may be used to obtain further information and to identify the tumor type.

In recent years an increasing amount of reports has appeared on biomarkers that may be used for risk assessment, screening, diagnosis, prognosis, and for selection and monitoring of therapies of lung cancer.

For example, Semmens et al, 2005 (WO 2005/098445) disclose protein biomarkers that are differentially present in the samples of patients with lung cancer and in the samples of control subjects and can thus be used in diagnosing lung cancer. The measurement of these markers, alone or in combination, in patient samples, is reported to provide information that can be correlated with a probable diagnosis of lung cancer or a negative diagnosis (e. g., normal or disease-free). All the markers are characterized by molecular weight.

Gold et al, 2011 (WO 2011/031344) disclose a list of 61 biomarkers that can be used alone or in various combinations to diagnose lung cancer.

YIP et al, 2004 (WO 2004/061410) disclose a method for qualifying lung carcinoma status in a subject, which comprises analyzing a biological sample from said subject for a diagnostic level of a biomarker protein, which is differentially present in samples of a subject with lung cancer and a normal subject that is free of lung cancer. YIP et al, 2004 provide a list of about 50 preferred biomarkers that may be used in such a method.

Sungwhan et al, 2007 (US 2007/0264659) disclose an epigenetic marker for lung cancer.

Cigarette smoke is a major risk factor for lung cancer however, there is a subset of patients that develop lung cancers despite no history of smoking. Lung cancer in smokers and non-smokers are biologically different, with the former being usually more aggressive and resistant to conventional therapies, which often results in poor outcome and decreased survival.

The underlying molecular mechanisms promoting lung cancer in smokers are still not well understood. Despite the increasing number of reports on biomarkers for lung cancer detection and classification, there is still a need for identification of biomarkers and biological pathways that are specific lung cancers in smokers, which may best be targeted with different treatments. This need could be met within the scope of the present invention.

The present invention provides biomarkers or a panel of biomarkers that are predictive for smoking-related lung cancer development and that can be used to develop tools in order to decrease adverse health effects caused by exposure to tobacco smoke and other tobacco-related products. In particular, the present invention provides biomarkers, which were identified based on gene expression patterns that are associated with smoking-related lung cancer development and can be used to provide a prognosis of a cancer or to predict the outcome of a treatment, or both.

The present invention relates to a method of providing a prognosis on the development of lung cancer in an individual, said method comprising the steps of

i. determining in a sample taken from an individual, whose prognosis on lung cancer is to be provided, a test value for the expression of a biomarker; and

ii. determining in a sample taken from a non-smoker suffering from lung cancer and/or from a smoker suffering from lung cancer a reference value for the expression of a biomarker, and

iii. comparing the test values obtained for each biomarker with the reference values obtained from a non-smoker suffering from lung cancer and/or a smoker suffering from lung cancer,

wherein said biomarker corresponds to at least 2 of the genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2)

and 213523_at (CCNE1), optionally, in combination with at least 1 and up to 108 genes, which are different from the above identified genes and are depicted depicted in table 1, table 2, or both table 1 and table 2.

In one embodiment, the present invention thus provides a method of providing a prognosis on the development of lung cancer in an individual, said method comprising the steps of

i. determining in a sample taken from an individual, whose prognosis on lung cancer is to be provided, a test value of the expression of a biomarker comprising using a composition, comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotidecorresponding to at least 2 and up to 111 different genes depicted in table 1, table 2, or both table 1 and table 2; and

ii. comparing the values obtained for each biomarker in the panel with reference values for the biomarkers from lung cancer of a non-smoker;

wherein differences in the values indicate a poor prognosis on the development of the lung cancer in the individual.

In one embodiment, the present invention thus provides a method of providing a prognosis on the development of lung cancer in an individual, said method comprising the steps of

i. determining in a sample taken from an individual, whose prognosis on lung cancer is to be provided, a test value of the expression of a biomarker comprising using a composition, comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide corresponding to at least 2 of the genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally in combination with at least 1 and up to 108 additional genes which are different from the above identified genes and are depicted in table 1, table 2, or both table 1 and table 2; and

ii. comparing the values obtained for each of the biomarkers used in step (i) with reference values for the same biomarkers from lung cancer of a non-smoker;

wherein differences in the values indicate a poor prognosis on the development of the lung cancer in the individual.

In one embodiment, the present invention thus provides a method of providing a prognosis on the development of lung cancer in an individual, said method comprising the steps of

i. determining in a sample taken from an individual, whose prognosis on lung cancer is to be provided, a test value of the expression of a biomarker comprising using a composition, comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide corresponding to the genes identified by GeneBank Accession numbers 204767_s_at (FEN1) and 209545_s_at (RIPK2); or 204767_s_at (FEN1) and 213523_at (CCNE1); or 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally in combination with at least 1 and up to 108 genes, which are different from the above identified genes and are depicted in table 1, table 2, or both table 1 and table 2; and

ii. comparing the values obtained for each biomarker in the panel with reference values for the biomarkers from lung cancer of a non-smoker;

wherein differences in the values indicate a poor prognosis on the development of the lung cancer in the individual.

In one embodiment, the present invention thus provides a method of providing a prognosis on the development of lung cancer in an individual, said method comprising the steps of

i. determining in a sample taken from an individual, whose prognosis on lung cancer is to be provided, a test value of the expression of a biomarker comprising using a composition, comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide corresponding to the genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally in combination with at least 1 and up to 108 genes, which are different from the above identified genes and are depicted in table 1, table 2, or both table 1 and table 2; and

ii. comparing the values obtained for each biomarker in the panel with reference values for the biomarkers from lung cancer of a non-smoker;

wherein differences in the values indicate a poor prognosis on the development of the lung cancer in the individual.

In a specific embodiment, at least one of the biomarker genes is a gene depicted in table 3.

Differential expression may be determined by way of comparison to a control sample, particularly a control sample obtained from an individual, which is a non-smoker and suffers from lung cancer. Alternatively, differential expression can also be determined by comparison with a reference value, wherein the value is obtained from a non-smoker who has lung cancer or from a population of non-smokers who have lung cancer. The difference in gene expression can be represented by the value of fold change or a logarithm of the value of the fold change in base 2 (herein referred to as log fold change).

In various embodiments of the invention, the difference in expression of the at least two, and up to 111, biomarkers in the test sample is determined to have changed relative to that in the control sample by a log fold change value of more than 0.35, 0.4, 0.5, 0.6, 0.7 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.25, 2.5, 2.75, 3.0, 3.25, 3.5, 3.75, 4.0, 4.25, 4.5, 4.75, or 5.0.

In various embodiments of the invention, the difference in expression of the at least 2 of the genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally in combination with at least 1 and up to 108 genes, which are different from the above identified genes and are depicted in table 1, table 2, or both table 1 and table 2 in the test sample, is determined to have changed relative to that in the control sample by a log fold change value of more than 0.35, 0.4, 0.5, 0.6, 0.7 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.25, 2.5, 2.75, 3.0, 3.25, 3.5, 3.75, 4.0, 4.25, 4.5, 4.75, or 5.0.

In certain embodiments of the invention, the sample is obtained from the respiratory system, such as but not limited to lung cells.

In certain embodiments of the invention, the genes representing the biomarkers are differentially expressed in lung adenocarcinomas from smokers when compared to

adenocarcinomas from non-smokers and are involved in one or more of the following networks.

a. modulation of the immune response in terms of immune evasion of the tumor;
b. apoptosis and cell survival network; and/or
c. cellular growth and proliferation network.

In one embodiment, the present invention relates to a method of any one of the preceding embodiments, comprising

i. determining in a biological sample of an individual who suffers from lung cancer a test value for the expression of each of the at least 2 and up to 111 different biomarker genes depicted in table 1, table 2, or both table 1 and table 2;
ii. determining in a biological sample of a control individual a reference value for the expression of each of the at least 2 and up to 111 different biomarker genes depicted in table 1, table 2, or both table 1 and table 2; and
iii. comparing the values obtained for each said biomarker genes with the reference value;

wherein said control individual is a non-smoker suffering from lung cancer and wherein differences between the test values and the reference values indicate a prognosis on the outcome of the cancer treatment.

In one embodiment, the present invention relates to a method of any one of the preceding embodiments, comprising

i. determining in a biological sample of an individual who suffers from lung cancer a test value for the expression of each of the at least 2 of the biomarker genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally in combination with at least 1 and up to 108, which are different from the above identified genes and are depicted in table 1, table 2, or both table 1 and table 2;
ii. determining in a biological sample of a control individual a reference value for the expression of each of the same biomarker genes used in step (i); and
iii. comparing the values obtained for each said biomarker genes with the reference value;

wherein said control individual is a non-smoker suffering from lung cancer and wherein differences between the test values and the reference values indicate a prognosis on the outcome of the cancer treatment.

In one embodiment, the present invention relates to a method of any one of the preceding embodiments, comprising

i. determining in a biological sample of an individual who suffers from lung cancer a test value for the expression of the genes identified by GeneBank Accession numbers 204767_s_at (FEN1) and 209545_s_at (RIPK2); or 204767_s_at (FEN1) and 213523_at (CCNE1); or 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally in combination with at least 1 and up to 108 genes, which are different from the above identified genes and are depicted in table 1, table 2, or both table 1 and table 2;
ii. determining a reference value in a sample taken from a non-smoker suffering from lung cancer and/or a smoker suffering from lung cancer comprising using a composition comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide corresponding to the genes identified in section i); andcomparing the values obtained for each said biomarker genes with the reference value;

wherein said control individual is a non-smoker suffering from lung cancer and wherein differences between the test values and the reference values indicate a prognosis on the outcome of the cancer treatment.

In one embodiment, the present invention relates to a method of any one of the preceding embodiments, comprising

i. determining in a biological sample of an individual who suffers from lung cancer a test value for the expression of each of the genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally in combination with at least 1 and up to 108 genes, which are different from the above identified genes and are depicted in table 1, table 2, or both table 1 and table 2;
ii. determining a reference value in a sample taken from a non-smoker suffering from lung cancer and/or a smoker suffering from lung cancer comprising using a composition comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide

corresponding to the genes identified in section i); andcomparing the values obtained for each said biomarker genes with the reference value;

wherein said control individual is a non-smoker suffering from lung cancer and wherein differences between the test values and the reference values indicate a prognosis on the outcome of the cancer treatment.

The difference between or the similarity in the test values and the reference values may be used to forecast or classify the subject into a poor survival group or a good survival group.

The invention further provides a method for predicting or monitoring the outcome of a lung cancer treatment in an individual suffering from lung cancer comprising

i. determining in a biological sample of an individual who suffers from lung cancer a value for the expression of each of the at least 2 and up to 111 different biomarker genes depicted in table 1, table 2, or both table 1 and table 2;
ii. determining a reference value in a sample taken from a non-smoker suffering from lung cancer and/or a smoker suffering from lung cancer comprising using a composition comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide corresponding to the genes identified in section i); andand comparing the values obtained for each said biomarker genes with the reference value;

wherein said control individual is a non-smoker suffering from lung cancer, wherein at least one of the biomarker genes is a gene depicted in table 3 and wherein differences between the test values and the reference values indicate the probability of said individual's lung cancer sharing one or more treatment outcomes of smoking-related lung cancer.

The invention further provides a method for predicting or monitoring the outcome of a lung cancer treatment in an individual suffering from lung cancer comprising

i. determining in a biological sample of an individual who suffers from lung cancer a value for the expression of each of the at least 2 of the genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally in combination

with at least 1 and up to 108 genes, which are different from the above identified genes and are depicted in table 1, table 2, or both table 1 and table 2;

ii. determining in a biological sample of a control individual a reference value for the expression of each of the same biomarker genes used in step (i); and comparing the values obtained for each said biomarker genes with the reference value;

wherein said control individual is a non-smoker suffering from lung cancer, wherein differences between the test values and the reference values indicate the probability of said individual's lung cancer sharing one or more treatment outcomes of smoking-related lung cancer.

In various embodiments of the invention, the difference between a test value and a reference value which indicates the development of the lung cancer or provides a prognosis is a log fold change value of more than 0.35, 0.4, 0.5, 0.6, 0.7 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.25, 2.5, 2.75, 3.0, 3.25, 3.5, 3.75, 4.0, 4.25, or 4.5.

For a biomarker gene which is up-regulated in lung cancer of smokers (table 1 genes), a positive log fold change value of between 0.35 and 5.0 indicates a poor prognosis on the development of the lung cancer in the tested individual, and *vice versa*.

Accordingly, for a biomarker gene which is down-regulated in lung cancer of smokers (table 2 genes), a negative log fold change value of between 0.35 and 5.0 also indicates a poor prognosis on the development of the lung cancer in the tested individual, and *vice versa*.

The log fold change value of each of the biomarker genes in table 1 and 2 can be used as exemplary threshold for each of the biomarker gene.

The samples used in the method according to the invention as described herein may be blood, serum, plasma, sputum, saliva, tissue particularly lung tissue, obtained through biopsy, bronchia brushings, exhaled breath, or urine.

In a specific embodiment of the invention, a biomarker or a panel of biomarkers may be used in the method according to the invention as described herein, which is

(a) selected from the group of biomarker genes depicted in table 1 identified by GeneBank Accession number 200887_s_at (STAT1), 204785_x_at (IFNAR2); 201762_s_at (PSME2), 204279_at (PSMB9), 203915_at (CXCL9), 204533_at (CXCL10), 227458_at (CD274), 207375_s_at (IL15), 207375_s_at (IL15RA), 205207_at (IL6), 204655_at (CCL5), 204103_at (CCL4), 207072_at (IL18RAP), and 205291_at (IL2RB); or

(b) selected from the group of biomarker genes depicted in table 1 identified by GeneBank Accession number 203213_at (CDK1), 204767_s_at (FEN1), 213523_at (CCNE1), 222036_s_at (MCM4), 202870_s_at (CDC20); 202095_s_at (BIRC5), 202246_s_at (CDK4), and 207574_s_at (GADD45B); or

(c) selected from the group of biomarker genes depicted in table 1 and table 2 identified by GeneBank Accession number 208373_s_at (P2RY6), 210563_x_at (CFLAR), 210538_s_at (BIRC3), 207535_s_at (NFKB2), 203868_s_at (CHEK1), 203666_at (CXCL12), 209875_s_at (SPP1), 207697_x_at (LILRB2), 204639_at (ADA), 209545_s_at (RIPK2), and 224818_at (SORT1).

In a specific embodiment of the invention, a biomarker or a panel may be used in the method according to the invention as described herein of at least 2 of the genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally in combination with at least 1 biomarker gene, which is

(a) selected from the group of biomarker genes depicted in table 1 identified by GeneBank Accession number 200887_s_at (STAT1), 204785_x_at (IFNAR2); 201762_s_at (PSME2), 204279_at (PSMB9), 203915_at (CXCL9), 204533_at (CXCL10), 227458_at (CD274), 207375_s_at (IL15), 207375_s_at (IL15RA), 205207_at (IL6), 204655_at (CCL5), 204103_at (CCL4), 207072_at (IL18RAP), and 205291_at (IL2RB); or

(b) selected from the group of biomarker genes depicted in table 1 identified by GeneBank Accession number 203213_at (CDK1), 222036_s_at (MCM4), 202870_s_at (CDC20); 202095_s_at (BIRC5), 202246_s_at (CDK4), and 207574_s_at (GADD45B); or

(c) selected from the group of biomarker genes depicted in table 1 and table 2 identified by GeneBank Accession number 208373_s_at (P2RY6),

210563_x_at (CFLAR), 210538_s_at (BIRC3), 207535_s_at (NFKB2), 203868_s_at (CHEK1), 203666_at (CXCL12), 209875_s_at (SPP1), 207697_x_at (LILRB2), 204639_at (ADA), and 224818_at (SORT1).

In a specific embodiment of the invention, a biomarker or a panel may be used in the method according to the invention as described herein of the genes identified by GeneBank Accession numbers 204767_s_at (FEN1) and 209545_s_at (RIPK2); or 204767_s_at (FEN1) and 213523_at (CCNE1); or 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally in combination with at least 1 biomarker gene, which is

(a) selected from the group of biomarker genes depicted in table 1 identified by GeneBank Accession number 200887_s_at (STAT1), 204785_x_at (IFNAR2); 201762_s_at (PSME2), 204279_at (PSMB9), 203915_at (CXCL9), 204533_at (CXCL10), 227458_at (CD274), 207375_s_at (IL15), 207375_s_at (IL15RA), 205207_at (IL6), 204655_at (CCL5), 204103_at (CCL4), 207072_at (IL18RAP), and 205291_at (IL2RB); or

(b) selected from the group of biomarker genes depicted in table 1 identified by GeneBank Accession number 203213_at (CDK1), 222036_s_at (MCM4), 202870_s_at (CDC20); 202095_s_at (BIRC5), 202246_s_at (CDK4), and 207574_s_at (GADD45B); or

(c) selected from the group of biomarker genes depicted in table 1 and table 2 identified by GeneBank Accession number 208373_s_at (P2RY6), 210563_x_at (CFLAR), 210538_s_at (BIRC3), 207535_s_at (NFKB2), 203868_s_at (CHEK1), 203666_at (CXCL12), 209875_s_at (SPP1), 207697_x_at (LILRB2), 204639_at (ADA), and 224818_at (SORT1).

In a specific embodiment of the invention, a biomarker or a panel may be used in the method according to the invention as described herein of the genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally in combination with at least 1 biomarker gene, which is

(a) selected from the group of biomarker genes depicted in table 1 identified by GeneBank Accession number 200887_s_at (STAT1), 204785_x_at (IFNAR2); 201762_s_at (PSME2), 204279_at (PSMB9), 203915_at (CXCL9), 204533_at (CXCL10), 227458_at (CD274), 207375_s_at (IL15),

207375_s_at (IL15RA), 205207_at (IL6), 204655_at (CCL5), 204103_at (CCL4), 207072_at (IL18RAP), and 205291_at (IL2RB); or

(b) selected from the group of biomarker genes depicted in table 1 identified by GeneBank Accession number 203213_at (CDK1), 222036_s_at (MCM4), 202870_s_at (CDC20); 202095_s_at (BIRC5), 202246_s_at (CDK4), and 207574_s_at (GADD45B); or

selected from the group of biomarker genes depicted in table 1 and table 2 identified by GeneBank Accession number 208373_s_at (P2RY6), 210563_x_at (CFLAR), 210538_s_at (BIRC3), 207535_s_at (NFKB2), 203868_s_at (CHEK1), 203666_at (CXCL12), 209875_s_at (SPP1), 207697_x_at (LILRB2), 204639_at (ADA), and 224818_at (SORT1).

In a specific embodiment of the invention, at least one of the biomarker genes in the above identified panel of biomarkers is a gene from table 3.

Fragments of the above identified genes or complementary sequences of said genes, particularly sequences having 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% sequence homology with said gene sequences may also be used within the method according to the invention as described herein in the various embodiments.

The biomarker genes identified in section a) above are up-regulated in lung tumors in smokers and are involved in cell-to-cell signalling and the modulation of the immune response. The overall effect of this group of genes is to promote immune evasion of the tumor.

The biomarker genes identified in section b) above are up-regulated in lung tumors in smokers. The overall effect of these genes is to promote tumor growth by up-regulating processes involved in cell proliferation: DNA replication, cell cycle progression, mitosis, as well the repair of DNA damage.

The biomarker genes identified in section c) above are up-regulated in lung tumors in smokers, with the exception of SORT1, which is down-regulated. The overall effect is to promote tumor cell survival and protect tumor cells from the activation of apoptosis and other cell death pathways.

In one embodiment of the invention, at least one of the biomarker genes selected from the group of genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1) may be used in the method according to the invention as described herein. In particular, at least two of the

biomarker genes selected from the group of genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1) may be used in the method according to the invention as described herein. In particular, the biomarker genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1) may be used in the method according to the invention as described herein. Especially, the above biomarkers may optionally be used in combination with at least 1 and up to 108 genes, which are different from the above identified genes and are depicted in table 1, table 2, or both table 1 and table 2.

In a further embodiment of the invention, at least one of the biomarker genes selected from the genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1) may optionally be used in combination with at least one of the biomarker genes depicted in table 1, table 2 and/or table 3 in the method according to the invention as described herein. In particular, at least two of the biomarker genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1) may optionally be used in combination with at least one of the biomarker genes depicted in table 1, table 2 and/or table 3 in the method according to the invention as described herein. In particular, the three biomarker genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1) may optionally be used in combination with at least one of the biomarker genes depicted in table 1, table 2 and/or table 3 in the method according to the invention as described herein.

In one embodiment, a polynucleotide or a variant thereof may be used in any one of the methods of the invention described herein in the various embodiments, which polynucleotide is complementary to a target gene as depicted in table 1, table 2, or both table 1 and table 2, and is used as a molecular probe in a hybridization reaction or as a molecular primer in a nucleic acid extension reaction, for the determination of the target and reference value. In a specific embodiment of the invention, at least one of the polynucleotides or a variant thereof is complementary to a biomarker gene from table 3.

In one embodiment, one or more detectably labeled antibodies may be used in any one of the methods of the invention described herein in the various embodiments for

the determination of the target and reference value, which antibodies are capable of identifying biomarker gene products encoded by one or more biomarker genes depicted in table 1, table 2, or both table 1 and table 2, or by conserved variants or peptide fragments thereof. In a specific embodiment of the invention, at least one of the detectably labeled antibodies is capable of identifying a biomarker gene product encoded by one or more biomarker genes depicted in table 3.

In certain embodiments of the invention, determination of biomarker values is accomplished by performing an *in-vitro* assay, particularly an *in-vitro* assay selected from the group consisting of an antibody-based assay such as an immunoassay, a histological or cytological assay, an expression level assay such as an RNA expression level assay and an aptamer-based assay.

In certain further embodiments of the invention, the biomarker value may be determined by performing

a. mass spectrometry;
b. an enzyme-linked immunosorbent assay;
c. a microarray-based immunohistochemical analysis; or
d. high-throughput DNA sequencing.

In a specific embodiment of the invention, the biomarker value is determined by performing mass spectrometry.

In another specific embodiment of the invention, an enzyme-linked immunosorbent assay may be performed for determining biomarker values.

In another specific embodiment of the invention, a microarray-based immunohistochemical analysis may be performed for determining biomarker values.

In still another specific embodiment of the invention, surface enhanced laser desorption/ionization may be performed for determining biomarker values.

In one aspect of the invention, data analysis is performed within a method according to any one of the preceding embodiments by one or more computer program(s).

In one embodiment, the present invention provides a panel of at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, or of all 111 biomarker genes selected from the group of biomarker genes provided in table 1, table 2, or

both table 1 and table 2, for use in a method according to any one of the preceding embodiments. In a specific embodiment of the invention, at least one of the biomarker genes in this panel is a gene from table 3.

In one embodiment, the present invention provides a panel of (a) at least 2, at least 3, at least 4, at least 5 biomarker genes selected from the group of markers depicted in table 3; (b) and at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50 biomarker genes selected from the group of markers depicted in table 1, table 2, or both table 1 and table 2 for use in a method according to the invention as described in any one of the preceding embodiments.

In a specific embodiment, a panel of at least 10 biomarker genes is provided selected from the group of markers depicted in table 1, table 2, or both table 1 and table 2 for use in a method according to any one of the preceding embodiments. In a specific embodiment of the invention, at least one of the biomarker genes in this panel is a gene from table 3.

In another specific embodiment of the invention, a panel of at least 50 biomarker genes is provided selected from the group of markers depicted in table 1, table 2, or both table 1 and table 2 for use in a method according to any one of the preceding embodiments. In a specific embodiment of the invention, at least one of the biomarker genes in this panel is a gene from table 3.

In one embodiment of the invention, the panel comprises a polynucleotide or a variant thereof, which is complementary to a target gene as depicted in table 1, table 2, or both table 1 and table 2 and can be used as a hybridization probe or a primer. In a specific embodiment of the invention, at least one of the polynucleotides or a variants thereof is complementary to a biomarker gene from table 3.

In one embodiment, the present invention provides a kit for predicting a prognosis on the outcome of cancer treatment in an individual suffering from lung cancer or for predicting or monitoring the outcome of a lung cancer treatment in such an individual, comprising a reagent for detecting differential expression of a panel of at least 2 and up to 50, particularly up to 111 different biomarkers selected from the biomarkers depicted in table 1, table 2, or both table 1 and table 2. In a specific embodiment of the invention, at least one of the biomarker genes in this panel is a gene from table 3.

In one embodiment, the present invention provides a kit for predicting a prognosis on the outcome of cancer treatment in an individual suffering from lung cancer or for

predicting or monitoring the outcome of a lung cancer treatment in such an individual, comprising a device or reagent for detecting differential expression of biomarkes comprising using a composition, in particular a panel of at least 2 of the genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally in combination with at least 1 and up to 108 genes, which are different from the above identified genes and are depicted in table 1, table 2, or both table 1 and table 2.

In one embodiment, the present invention provides a kit for predicting a prognosis on the outcome of cancer treatment in an individual suffering from lung cancer or for predicting or monitoring the outcome of a lung cancer treatment in such an individual, comprising a device or reagent for detecting differential expression of a panel of the genes identified by GeneBank Accession numbers 204767_s_at (FEN1) and 209545_s_at (RIPK2); or 204767_s_at (FEN1) and 213523_at (CCNE1); or 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally in combination with at least 1 and up to 108 genes, which are different from the above identified genes and are depicted in table 1, table 2, or both table 1 and table 2.

In one embodiment, the present invention provides a kit for predicting a prognosis on the outcome of cancer treatment in an individual suffering from lung cancer or for predicting or monitoring the outcome of a lung cancer treatment in such an individual, comprising a device or reagent for detecting differential expression of a panel of genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally in combination with at least 1 and up to 108 genes, which are different from the above identified genes and are depicted in table 1, table 2, or both table 1 and table 2.

In a specific aspect, the present invention relates to a panel of biomarkers selected from the biomarker genes depicted in table 1, table 2, or both table 1 and table 2 comprising those biomarker genes from said tables which are identified by GeneBank Accession identifiers and to the use of said biomarkers or panel of biomarkers in a method or a kit according to the invention and described herein in the various embodiments. In a specific embodiment of the invention, at least one of the biomarker genes in this panel is a gene from table 3.

In a further specific aspect, the present invention relates to a composition, particularly a panel of genes, comprising at least 2 different isolated nucleic acid molecules each

of which comprising a biomarker polynucleotide that corresponds to a different gene identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally further comprising at least 1 and up to 108 different nucleic acid molecules each of which comprises a biomarker polynucleotides corresponding to genes depicted in table 1, table 2, or both table 1 and table 2 and to the use of said nucleic acid molecules in a method or a kit according to the invention and described herein in the various embodiments.

In a further specific aspect, the present invention relates to a composition, particularly a panel of genes, comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide that corresponds to a different gene identified by GeneBank Accession numbers 204767_s_at (FEN1) and 209545_s_at (RIPK2); or 204767_s_at (FEN1) and 213523_at (CCNE1); or 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally in combination with at least 1 and up to 108 genes, which are different from the above identified genes and are depicted in table 1, table 2, or both table 1 and table 2 and to the use of said biomarkers or a composition, particularly a panel of biomarkers in a method or a kit according to the invention and described herein in the various embodiments.

In a further specific aspect, the present invention relates to a composition, particularly a panel of genes, comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide that corresponds to a different gene identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1) optionally in combination with at least 1 and up to 108 genes, which are different from the above identified genes and are depicted in table 1, table 2, or both table 1 and table 2 and to the use of said biomarkers or a composition, particularly a panel of biomarkers in a method or a kit according to the invention and described herein in the various embodiments.

In certain embodiments, the present invention relates to a composition, particularly a panel of genes, comprising comprising genes

(a) selected from the group of biomarker genes depicted in table 1 identified by GeneBank Accession number 200887_s_at (STAT1), 204785_x_at (IFNAR2); 201762_s_at (PSME2), 204279_at (PSMB9), 203915_at (CXCL9), 204533_at (CXCL10), 227458_at (CD274), 207375_s_at (IL15),

207375_s_at (IL15RA), 205207_at (IL6), 204655_at (CCL5), 204103_at (CCL4), 207072_at (IL18RAP), and 205291_at (IL2RB); or

(b) selected from the group of biomarker genes depicted in table 1 identified by GeneBank Accession number 203213_at (CDK1), 204767_s_at (FEN1), 213523_at (CCNE1), 222036_s_at (MCM4), 202870_s_at (CDC20); 202095_s_at (BIRC5), 202246_s_at (CDK4), and 207574_s_at (GADD45B); or

(c) selected from the group of biomarker genes depicted in table 1 and table 2 identified by GeneBank Accession number 208373_s_at (P2RY6), 210563_x_at (CFLAR), 210538_s_at (BIRC3), 207535_s_at (NFKB2), 203868_s_at (CHEK1), 203666_at (CXCL12), 209875_s_at (SPP1), 207697_x_at (LILRB2), 204639_at (ADA), 209545_s_at (RIPK2), and 224818_at (SORT1).

In a specific embodiment of the invention, at least one of the biomarker genes in this composition, particularly this panel is a gene from table 3.

In certain embodiments, the present invention relates to a composition, particularly a panel of genes, comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide that corresponds to a different gene comprising genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1).

In certain embodiments, the present invention relates to a composition, particularly a panel of genes, comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide that corresponds to a different gene comprising genes identified by GeneBank Accession numbers 204767_s_at (FEN1) and 209545_s_at (RIPK2).

In certain embodiments, the present invention relates to a composition, particularly a panel of genes, comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide that corresponds to a different gene comprising genes identified by GeneBank Accession numbers 204767_s_at (FEN1) and 213523_at (CCNE1).

In certain embodiments, the present invention relates to a composition, particularly a panel of genes, comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide that corresponds to a different gene

comprising genes identified by GeneBank Accession numbers 209545_s_at (RIPK2) and 213523_at (CCNE1).

In certain embodiments, the present invention relates to a composition, particularly a panel of genes, comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide that corresponds to a different gene comprising genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1) optionally in combination with at least one and up to 108 genes, which are different from the above identified genes and which are depicted in table 1, table 2, or both table 1 and table 2.

In certain embodiments, the present invention relates to a composition, particularly a panel of genes, comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide that corresponds to a different gene comprising genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally in combination with at least one and up to 108 genes, which are different from the above identified genes and which are depicted in table 1, table 2, and/or table 3.

In certain embodiments, the present invention relates to a composition, particularly a panel of genes, comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide that corresponds to a different gene of at least 2 of the biomarker genes comprising genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1).

In certain embodiments, the present invention relates to a composition, particularly a panel of genes, comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide that corresponds to a different gene of at least 2 of the biomarker genes comprising genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally in combination with at least one and up to 108 genes, which are different from the above identified genes and are depicted in table 1, table 2, or both table 1 and table 2.

In certain embodiments, the present invention relates to a composition, particularly a panel of genes, comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide that corresponds to a different gene of

at least 2 of the biomarker genes comprising genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally in combination with at least one and up to 108 genes, which are different from the above identified genes and are depicted in table 1, table 2, and/or table 3.

In certain embodiments, the present invention relates to a composition, particularly a panel of genes, comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide that corresponds to a different gene of the biomarker genes comprising genes identified by GeneBank Accession numbers 204767_s_at (FEN1) and 209545_s_at (RIPK2), or 204767_s_at (FEN1) and 213523_at (CCNE1), or 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally in combination with at least one of the biomarker genes

(a) selected from the group of biomarker genes depicted in table 1 identified by GeneBank Accession number 200887_s_at (STAT1), 204785_x_at (IFNAR2); 201762_s_at (PSME2), 204279_at (PSMB9), 203915_at (CXCL9), 204533_at (CXCL10), 227458_at (CD274), 207375_s_at (IL15), 207375_s_at (IL15RA), 205207_at (IL6), 204655_at (CCL5), 204103_at (CCL4), 207072_at (IL18RAP), and 205291_at (IL2RB); or

(b) selected from the group of biomarker genes depicted in table 1 identified by GeneBank Accession number 203213_at (CDK1), 222036_s_at (MCM4), 202870_s_at (CDC20); 202095_s_at (BIRC5), 202246_s_at (CDK4), and 207574_s_at (GADD45B); or

(c) selected from the group of biomarker genes depicted in table 1 and table 2 identified by GeneBank Accession number 208373_s_at (P2RY6), 210563_x_at (CFLAR), 210538_s_at (BIRC3), 207535_s_at (NFKB2), 203868_s_at (CHEK1), 203666_at (CXCL12), 209875_s_at (SPP1), 207697_x_at (LILRB2), 204639_at (ADA) and 224818_at (SORT1).

The term "composition" as used herein can be a mixtue of nucleic acid molecules in e.g. a solution or a microarray, wherein the nucleic acid molecule may be immobilized on a substrate. In particular, a composition may be a panel of genes.

Fragments of the above identified genes or complementary sequences of said genes, particularly sequences having 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% sequence homology with said gene sequences, may also be used within the

composition, particularly the panel of biomarkers according to the invention as described herein in the various embodiments.

The present invention provides biomarkers or a combination of biomarkers that are predictive for smoking-related lung cancer development and that can be used to develop tools for monitoring adverse health effects caused by exposure to tobacco smoke and other tobacco-related products.

In one aspect of the present invention, biomarkers (genes) were identified based on gene expression patterns that can detect smoking-related lung cancer development.

In another aspect, the present invention provides biomarkers that can be used to monitor the progress of a treatment or predict the outcome of a treatment, in an individual patient.

The present invention provides a method for predicting or assessing prognosis on the outcome of cancer treatment in an individual suffering from lung cancer, said method, comprising

i. determining in a biological sample of an individual who suffers from lung cancer a value for the expression of each of the at least 2 and up to 111 different biomarker genes depicted in table 1, table 2, or both table 1 and table 2;
ii. determining a reference value in a sample taken from a non-smoker suffering from lung cancer and/or a smoker suffering from lung cancer comprising using a composition comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide corresponding to the genes identified in section i); andd comparing the values obtained for each said biomarker genes with the reference value;

wherein said control individual is a non-smoker suffering from lung cancer and wherein differences between the test values and the reference values indicate a prognosis on the outcome of cancer treatment.

The biomarker genes depicted in table 1 are genes which have been found to be up-regulated in lung cancer cells of smokers, whereas the genes in table 2 are down-regulated.

According to the invention, a high level of expression of a biomarker gene of table 1 relative to a reference level determined in a control individual, who is a non-smoker

suffering from lung cancer, indicates a poor prognosis for the outcome of cancer treatment and thus a decreased survival duration relative to the control; whereas a lower level of expression of a biomarker gene, which is close to or below the reference level indicates an improved prognosis for the outcome of cancer treatment, which means a survival duration similar to or increased relative the control.

For the biomarker genes of table 2, a poor prognosis correlates with a low level of expression as compared to the reference level and an improved prognosis with an expression level which is similar to or above the reference level.

The reference level of a biomarker can be established from cells from characterized cell lines, or cell samples from a non-smoker who has lung adenocarcinoma or a population of non-smokers who have lung adenocarcinoma.

In one embodiment, the present invention provides a method for monitoring the progress of a lung cancer treatment in an individual, said method comprising determining at suitable time intervals before, during, or after lung cancer therapy, particularly at different time points during the treatment, in a sample taken from said individual differential expression of a composition, particularly a panel of at least 2 and up to 111 different biomarkers selected from the biomarkers depicted in Table 1, Table 2, or both Table 1 and Table 2.

In a specific embodiment of the invention, at least one of the biomarker genes in the composition, particularly the panel is a gene from table 3.

In particular, said method comprises

i. determining in a plurality of biological samples a value for the expression of each of the at least 2 and up to 111 different biomarker genes depicted in table 1, table 2, or both table 1 and table 2, wherein said plurality of biological samples are obtained at a plurality of time points from an individual having lung cancer and receiving a lung cancer therapy;
ii. comparing the values obtained for each said biomarker genes with values of each said biomarker gene obtained at another time point and optionally with reference values for each said biomarker gene of a non-smoker suffering from lung cancer;

wherein differences between the test values obtained at the plurality of time points, or changes in the differences between the test values and the reference values provide a prediction of the outcome of the lung cancer therapy.

The method of the invention comprises measuring at suitable time intervals before, during, or after lung cancer therapy, the amount of biomarker gene product. Any change or absence of change in the amount of the biomarker gene product can be identified and correlated with the effect of the treatment on the subject. In a preferred aspect, the method comprises determining the levels of biomarker gene product levels at different time points and to compare these values with a reference level. The observed changes in the differences between the test values and the reference values over time can then be correlated with the disease course, treatment outcome or overall survival.

In one embodiment of the present invention the Matthews Correlation Coefficient (MCC) for the prediction of the smoking status ranges between 0.2 and 1, particularly between 0.3 and 1, more particularly between 0.37 and 1.

Detection of the protein biomarkers described herein in a test sample may be performed in a variety of ways well known to those skilled in the art.

In one aspect, the methods of the invention rely on the detection of the presence or absence of biomarker gene expression, or the qualitative or quantitative assessment of either over- or under-expression of biomarker gene in a population of test cells relative to a standard. Such methods utilize reagents such as biomarker polynucleotides and biomarker antibodies as described herein.

In particular, the level of expression of a biomarker gene may be determined by measuring the amount of biomarker messenger RNA, for example, by DNA-DNA hybridization, RNA-DNA hybridization, reverse transcription-polymerase chain reaction (PCR), or real time quantitative PCR; followed by comparing the results to a reference based on samples from clinically-characterized patients and/or cell lines of a known genotype/phenotype. As an alternative to amplification techniques, hybridization assays can be performed.

For example, these techniques find application in microarray-based assays that can be used to detect and quantify the amount of biomarker gene transcript using cDNA- or oligonucleotide-based arrays. Microarray technology allows multiple biomarker gene transcripts and/or samples from different subjects to be analyzed in one reaction. Typically, mRNA isolated from a sample is converted into labeled nucleic acids by reverse transcription and optionally in vitro transcription (cDNAs or cRNAs labeled with, for example, Cy3 or Cy5 dyes) and hybridized in parallel to probes

present on an array. See, for example, Schulze et al., Nature Cell Biol., 3 (2001), E190; and Klein et al., J Exp Med, 2001, 1625-1638, which are incorporated herein by reference in their entirety. Standard Northern analyses can be performed if a sufficient quantity of the test cells can be obtained. Utilizing such techniques, quantitative as well as size related differences between biomarker transcripts can also be detected.

High-throughput, massively parallel DNA sequencing techniques, also referred to next generation sequencing techniques, may also be applied to measure the level of expression of a biomarker gene, e.g. sequencing with bridge amplification (Illumina), pyrosequencing (Roche Diagnostics), ligation-based methods, ion torrent technology (Life Sciences) and single-molecule sequencing (Pacific Bio). Many such methods are well known in the art and are automated in commercially available sequencing machines.

The present invention provides isolated biomarker polynucleotides or variants thereof, which can be used, for example, as hybridization probes or primers ("biomarker probes" or "biomarker primers") to detect or amplify nucleic acids encoding a biomarker polypeptide, particularly a biomarker polypeptide encoded by a biomarker gene or polynucleotide selected from the group depicted in table 1, table 2, or both table 1 and table 2.

Nucleic acid molecules comprising nucleic acid sequences encoding the biomarker polypeptides or proteins of the invention, or genomic nucleic acid sequences from the biomarker genes (e.g., intron sequences, 5' and 3' untranslated sequences), or their complements thereof (i.e., antisense polynucleotides), are collectively referred to as "biomarker genes", "biomarker polynucleotides" or "biomarker nucleic acid sequences" of the invention. The present invention also provides isolated biomarker polynucleotides or variants thereof, which can be used, for example, as hybridization probes or primers ("biomarker probes" or "biomarker primers") to detect or amplify nucleic acids encoding a polypeptide of the invention. The term "biomarker gene product" thus encompasses both mRNA as well as translated polypeptide as a gene product of a biomarker.

The isolated biomarker polynucleotide according to the invention may comprise flanking sequences (*i.e.*, sequences located at the 5' or 3' ends of the nucleic acid), which naturally flank the nucleic acid sequence in the genomic DNA of the organism

from which the nucleic acid is derived. However, an isolated polynucleotide does not include an isolated chromosome, and does not include the poly(A) tail of an mRNA, if present. For example, in various embodiments, the isolated biomarker polynucleotide can comprise less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the coding sequence in genomic DNA of the cell from which the nucleic acid is derived. In other embodiments, the isolated biomarker polynucleotide is about 10-20, 21-50, 51-100, 101-200, 201-400, 401-750, 751-1000, 1001-1500 bases in length.

In various embodiments, the biomarker polynucleotides of the invention are used as molecular probes in hybridization reactions or as molecular primers in nucleic acid extension reactions as described herein. In these instances, the biomarker polynucleotides may be referred to as biomarker probes and biomarker primers, respectively, and the biomarker polynucleotides present in a sample which are to be detected and/or quantified are referred to as biomarker polynucleotides. Two biomarker primers are commonly used in DNA amplification reactions and they are referred to as biomarker forward primer and biomarker reverse primer depending on their 5' to 3' orientation relative to the direction of transcription. A biomarker probe or a biomarker primer is typically an oligonucleotide which binds through complementary base pairing to a subsequence of a biomarker polynucleotide. The biomarker probe may be, for example, a DNA fragment prepared by amplification methods such as by PCR or it may be chemically synthesized. A double stranded fragment may then be obtained, if desired, by annealing the chemically synthesized single strands together under appropriate conditions or by synthesizing the complementary strand using DNA polymerase with an appropriate primer. Where a specific nucleic acid sequence is given, it is understood that the complementary strand is also identified and included as the complementary strand will work equally well in situations where the target is a double stranded nucleic acid. A nucleic acid probe is complementary to a target nucleic acid when it will anneal only to a single desired position on that target nucleic acid under proper annealing conditions which depend, for example, upon a probe's length, base composition, and the number of mismatches and their position on the probe, and must often be determined empirically. Such conditions can be determined by those of skill in the art.

In one aspect of the invention, biomarkers may be detected in the test sample by gene expression profiling. In these methods, mRNA is prepared from a sample and

mRNA expression levels are measured by reverse transcription quantitative polymerase chain reaction (RT-PCR followed with qPCR). RT-PCR is used to create a cDNA from the corresponding mRNA. The cDNA may be used in a qPCR assay to produce fluorescence as the DNA amplification process progresses. By comparison to a standard curve, qPCR can produce an absolute measurement such as number of copies of mRNA per cell. Northern blots, microarrays, Invader assays, and RT-PCR combined with capillary electrophoresis may be used to measure expression levels of mRNA in a sample. Further details are provided, for example, in "Gene Expression Profiling: Methods and Protocols", Richard A. Shimkets, editor, Humana Press, 2004 and US patent application 20100070191.

In another aspect of the invention, detection of the biomarkers described herein may also be accomplished by an immunoassay procedure. The immunoassay typically includes contacting a test sample with an antibody that specifically binds to or otherwise recognizes a biomarker, and detecting the presence of the antibody/biomarker complex in the sample. The immunoassay procedure may be selected from a wide variety of immunoassay procedures known to those skilled in the art such as, for example, competitive or non-competitive enzyme-based immunoassays, enzyme-linked immunosorbent assays (ELISA), radioimmunoassay (RIA), and Western blots, etc. Further, multiplex assays may be used, including antibody arrays, wherein several desired antibodies are placed on a support, such as a glass bead or plate, and reacted or otherwise contacted with the test sample.

Antibodies used in these assays may be monoclonal or polyclonal, and may be of any type such as IgG, IgM, IgA, IgD and IgE. Monoclonal antibodies may be used to bind to a specific epitope offered by the biomarker molecule, and therefore may provide a more specific and accurate result. Antibodies may be produced by immunizing animals such as rats, mice, and rabbits. The antigen used for immunization may be isolated from the samples or synthesized by recombinant protein technology. Methods of producing antibodies and of performing antibody-based assays are well-known to the skilled artisan and are described, for example, more thoroughly in Antibodies: A Laboratory Manual (1988) by Harlow & Lane; Immunoassays: A Practical Approach, Oxford University Press, Gosling, J. P. (ed.) (2001) and/or Current Protocols in Molecular Biology (Ausubel et al.) which is regularly and periodically updated.

In certain embodiments, the present invention also provides "biomarker antibodies" including polyclonal, monoclonal, or recombinant antibodies, and fragments and variants thereof, that immunospecifically binds the respective biomarker proteins or polypeptides encoded by the genes or cDNAs (including polypeptides encoded by mRNA splice variants) as listed in tables 1 and 2.

Various chemical or biochemical derivatives of the antibodies or antibody fragments of the present invention can be produced using known methods. One type of derivative which is diagnostically useful as an immunoconjugate comprising an antibody molecule, or an antigen-binding fragment thereof, to which is conjugated a detectable label. However, in many embodiments, the biomarker antibody is not labeled but in the course of an assay, it becomes indirectly labeled by binding to or being bound by another molecule that is labeled. The invention encompasses molecular complexes comprising a biomarker antibody and a label, as well as immunocomplexes comprising a biomarker polypeptide, a biomarker antibody, and immunocomplexes comprising a biomarker polypeptide, a biomarker antibody, and a label.

Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferones, fluoresceins, fluorescein isothiocyanate, rhodamines, dichlorotriazinylamine fluorescein, dansyl chloride, phycoerythrins, Alexa Fluor 647, Alexa Fluor 680, $DiIC_{19}(3)$, Rhodamine Red-X, Alexa Fluor 660, Alexa Fluor 546, Texas Red, YOYO-1 + DNA, tetramethylrhodamine, Alexa Fluor 594, BODIPY FL, Alexa Fluor 488, Fluorescein, BODIPY TR, BODIPY TMR, carboxy SNARF-1, FM 1-43, Fura-2, Indo-1, Cascade Blue, NBD, DAPI, Alexa Fluor 350, aminomethylcoumarin, Lucifer yellow, Propidium iodide, or dansylamide; an example of a luminescent material includes luminol; examples of bioluminescent materials include green fluorescent proteins, modified green fluorescent proteins, luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

Immunoassays for biomarker polypeptides will typically comprise incubating a sample, such as a biological fluid, a tissue extract, freshly harvested cells, or lysates of cells, in the presence of a detectably labeled antibody capable of identifying biomarker gene products or conserved variants or peptide fragments thereof, and detecting the bound antibody by any of a number of techniques well-known in the art. One way of measuring the level of biomarker polypeptide with a specific biomarker antibody of the present invention is by enzyme immunoassay (EIA) such as an enzyme-linked immunosorbent assay (ELISA) (Voller, A. *et al.*, *J. Clin. Pathol. 31*:507-520 (1978); Butler, J.E., *Meth. Enzymol. 73*:482-523 (1981); Maggio, E. (ed.), *Enzyme Immunoassay*, CRC Press, Boca Raton, FL, 1980). The enzyme, either conjugated to the antibody or to a binding partner for the antibody, when later exposed to an appropriate substrate, will react with the substrate in such a manner as to produce a chemical moiety which can be detected, for example, by spectrophotometric, or fluorimetric means.

The biological sample may be brought in contact with and immobilized onto a solid phase support or carrier such as nitrocellulose, or other solid support which is capable of immobilizing cells, cell particles or soluble proteins. The support may then be washed with suitable buffers followed by treatment with the detectably labeled biomarker antibody. The solid phase support may then be washed with the buffer a second time to remove unbound antibody. The amount of bound label on solid support may then be detected by conventional means. A well known example of such a technique is Western blotting.

In various embodiments, the present invention provides compositions comprising labeled biomarker polynucleotides, or labeled biomarker antibodies to the biomarker proteins or polypeptides or labeled biomarker polynucleotides and labeled biomarker antibodies to the biomarker proteins or polypeptides according to the invention as described herein.

In addition to antibody-based techniques, the biomarkers described herein may also be detected and quantified by mass spectrometry. Mass spectrometry is a method that employs a mass spectrometer to detect ionized protein markers or ionized peptides as digested from the protein markers by measuring the mass-to-charge ratio (m/z). Labeling of biomarkers (along with other proteins) with stable heavy isotopes (deuterium, carbon-13, nitrogen-15, and oxygen-18) can be used in quantitative

proteomics. These are either incorporated metabolically in sample cells cultured briefly in vitro, or directly in samples by chemical or enzymatic reactions. Light, and heavy labeled biomarker peptide ions segregate and their intensity values are used for quantification. For example, analytes may be introduced into an inlet system of the mass spectrometer and ionized in an ionization source, such as a laser, fast atom bombardment, plasma or other suitable ionization sources known to the art. The generated ions are typically collected by an ion optic assembly and introduced into mass analyzers for mass separation before their masses are measured by a detector. The detector then translates information obtained from the detected ions into mass-to-charge ratios.

The invention also provides compositions comprising biomarker polynucleotides, biomarker polypeptides, or biomarker antibodies according to the invention as described herein in the various embodiments. The invention further provides diagnostic reagents for use in the methods of the invention, such as but not limited to reagents for flow cytometry and/or immunoassays that comprise fluorochrome-labeled antibodies that bind to one of the biomarker polypeptides of the invention as described herein.

In one embodiment, the invention provides diagnostic reagents that comprise one or more biomarker probes, or one or more biomarker primers. A diagnostic reagent may comprise biomarker probes and/or biomarker primers from the same biomarker gene or from multiple biomarker genes. In another embodiment, the invention also provides diagnostic compositions that comprise one or more biomarker probes and biomarker polynucleotides, or one or more biomarker primers and target polynucleotides, or biomarker primers, biomarker probes and biomarker target polynucleotides.

In some embodiments, the diagnostic compositions comprise biomarker probes and/or biomarker primers and a sample suspected to comprise biomarker target polynucleotides. Such diagnostic compositions comprise biomarker probes and/or biomarker primers and the nucleic acid molecules (including RNA, mRNA, cRNA, cDNA, and/or genomic DNA) of a subject in need of a diagnosis/prognosis of lung cancer.

The present invention also provides kits for practicing the methods of the invention. The kits can be used for clinical diagnosis and/or laboratory research. In one

embodiment, a kit comprises one or more diagnostic reagents in one or more containers. Preferably, the kit also comprises instructions in any tangible medium on the use of the diagnostic reagent(s) in one or more methods of the invention.

For nucleic acid-based methods, such as hybridization assays or polymerase chain reaction, a diagnostic reagent in the kit may comprise at least one biomarker polynucleotide, biomarker probe, and/or biomarker primer based on the biomarkers depicted in table 1, table 2, or both table 1 and table 2. The diagnostic reagents may be labeled, for example, by one or more different fluorochromes. Such a kit may optionally provide in separate containers enzymes and/or buffers for reverse transcription, in vitro transcription, and/or DNA polymerization, nucleotides, and/or labeled nucleotides, including fluorochrome-labeled nucleotides. Also included in the kit may be positive and negative controls for the methods of the invention.

For protein-based methods, such as immunoassays, a diagnostic reagent in the kit may comprise a biomarker antibody, which may be labeled, for example, by a fluorochrome. Such a kit may optionally provide in separate containers buffers, secondary antibodies, signal generating accessory molecules, labeled secondary antibodies, including fluorochrome-labeled secondary antibodies. The kit may also include unlabeled or labeled antibodies to various cell surface antigens which can used for identification or sorting of subpopulations of cells. Also included in the kit may be positive and negative controls for the methods of the invention. The positive and/or negative controls included in a kit can be nucleic acids, polypeptides, cell lysate, cell extract, whole cells from patients, or whole cells from cell lines.

The classification of the subject into the smoker-like group relates to the biomarker signature determined for the subject. The classification into smoker-like signature may be independent of the smoking status, that means also non-smoking subjects may be comprised, which reveal a biomarker gene signature, which is usually a typical smoker tumor gene signature as described herein. If a subject is classified into the smoker-like signature group according to the present invention, the subject has a poor survival prognosis.

Brief Description of the Figures

Figure 1: Interaction term. (A) Change in expression of this gene is due only to cigarette smoke. (B) Change in expression of this gene is specifically related to smoking-related lung cancers. Y-axis represents expression level of the gene.

Figure 2: Apoptosis and cell survival network that characterizes cancers from smokers. The numbers below the biomarkers are the log value of the fold change and the p value.Solid lines indicate a direct interaction between the biomarkers; dotted lines indicate an indirect interaction; loops indicate self regulation.

Figure 3: Volcano plot for the interaction model. The coefficients of the interaction model, gene=β_0+ β_1Smoking + β_2*Tissue+ β_3*TissuexSmoking+e, are estimated using LIMMA. Figure 3A - Volcano plot for the coefficients associated with the smoking effect (β_1), B) Volcano plot fort he coefficients associated with the tumor tissue effect (β_2) andC) Volcano plot for coefficients associated with the interaction effect (β_2). The genes that show significant interaction effect were selected for subsequent analysis as they capture the differences in gene expression between healthy and tumor tissue in smokers and non-smokers.

Figure 4: Biological processes and canonical functions associated with lung adenocarcinoma in smokers using Ingenuity Pathway Analysis. The top ten biological functions (A) and canonical pathways (B) are grouped based on p values (Fisher exact test). The threshold is less than 0.05.

Figure 5: Immune response and cell-to-cell signaling network. The molecules in the network represent genes upregulated in lung tumors from smokers when compared to both healthy tissue and tumors from non-smokers. Grey-shaded sections indicate groups of molecules participating in the processes indicated. Color intensity is a qualitative representation of fold-change. Straight lines indicate direct interaction. Dashed lines indicate indirect interactions.

Figure 6: Cell proliferation (A) and cell survival (B) subnetworks. The molecules in the network represent genes dysregulated in lung tumors from smokers when compared to both healthy tissue and tumors from non-smokers. Grey color indicates upregulated genes. Except from SORT1, which is a downregulated gene. White color indicates no significant differences in fold change. Color intensity is a qualitative representation of fold-change. Straight lines indicate direct interaction. Dashed lines indicate indirect interactions.

Figure 7: Interactions between all networks. Note how the upstream regulators IFNG, TNF and IL1B are heavily interacting with multiple components of the networks. Molecules represent genes dysregulated in lung tumors from smokers when compared to both healthy tissue and tumors from non-smokers. Grey color

indicates upregulated genes. Except from HEXB, CAT and SORT1, which are downregulated genes. White color indicates no significant differences in fold change. Color intensity is a qualitative representation of fold-change. Straight lines indicate direct interaction. Dashed lines indicate indirect interactions.

Figure 8: Heatmap of gene signature in training data. The three genes identified for deriving a predictive signature shows a differentiation between smokers and non-smokers in the tumors. These three genes have a higher average expression in smokers as compared to non-smokers. Furthermore this difference is specific to the tumour samples as those genes have a significant interaction effect. Grey and black colors in the side bar denote non-smokers and smokers respectively. Grey scales in the heatmap denote positive (unfilled) and negative (dashed lines) Z-scores respectively. Grey scale color intensity is a qualitative representation of fold-change.

EXAMPLES

The foregoing description will be more fully understood with reference to the following Examples. The Examples are, however, exemplary methods of practicing the present invention and are not intended to limit the scope of the invention. The present invention is based on a study to identify those genes that are affected differentially in tumor tissue and healthy tissue due to exposure to cigarette smoke.

1. Study design and subject selection

Tumor and normal lung tissue samples were obtained from 120 subjects, including smokers (n=60) and non-smokers (n=60) from Indivumed Biobank (Indivumed GmbH, Hamburg, Germany). Non-smokers are defined as either subjects that had never smoked during life or those that had stopped smoking for at least 16 years before samples were obtained. The rationale behind the incorporation of former smokers that have not smoked for 16 years is that the risk of developing lung cancer for this group is comparable to the risk of a person that has never smoked. Subject and sample selection was strictly based on the study purpose and the availability of tissue samples according to the experimental requirements, demographic and clinical data. The group composition was analyzed with Statistica software (Statsoft, Tulsa, US) which resulted in overall homogenous subject population. Ethical guidelines and confidentiality have been strictly assured and subjects gave written consent to participate in the study.

Normal lung tissue was obtained from smokers (n=30) and non-smokers (n=30) treated from non-cancerous lung diseases such as hamartoma, chronic pneumonia, granuloma or normal lung tissue from patients with lung metastasis. Lung tumor tissue was obtained from smokers (n=30) and non-smokers (n=30) with lung cancer histologically classified as adenocarcinoma of the lung. None of the adenocarcinoma patients had a metastasis.

2. Tissue collection and preparation

Tissue collection and preparation was performed at Indivumed Biobank. Freshly frozen lung tissue samples from the subject population were analyzed for normal and tumor cell content by staining before further sectioning. All tissues were collected under standardized conditions, snap frozen and stored in liquid nitrogen. The ischemia times are in all cases below 15 minutes. Each tissue block was quality controlled and only tissues with low amount of necrotic and / or apoptotic areas (<10%) as well as a tumor content of at least 40-50% were used for the study.

For the analysis, 5µM sections were cut in a cryostat and H&E stained. Stained sections were analyzed by microscopy. 16-32 x 20µM sections, depending on tissue size, were cut and re-suspended in 500µl RNA later solution (Qiagen) with RNAse inhibitor (Applied Biosystems) to assure RNA integrity and stored at -80°C until extraction.

3. RNA extraction

Frozen tissue sections stored in RNA later / RNAse inhibitor solution were homogenized in RLT lysis buffer (Qiagen). RNA samples were DNAse-treated and extracted with Qiagen RNeasy kit following manufacturer's instructions. The quality of the prepared RNA was documented by microcapillary electrophoresis using an Agilent 2100 Bioanalyzer (Agilent Technologies). Only those RNA samples with a RIN number ≥7.0 were selected for hybridization.

4. cDNA synthesis and amplification

cDNA and cRNA were synthetized using the One Cycle cDNA kit (Affymetrix) and following manufacturer's instructions. Unincorporated nucleotides were removed by using Qiagen RNeasy purification columns (Qiagen). The RNA yield was determined by measuring the absorption at 260nm.

5. Fragmentation, hybridization, detection

Fragmentation of the cRNA was performed in 2X fragmentation buffer (Affymetrix kit component). The fragmented cRNA was examined with an Agilent 2100 Bioanalyzer (Agilent Technologies) using non-fragmented cRNA as a control. Each fragmented cRNA sample was hybridized onto Affymetrix HG U133 plus 2.0 chips together with hybridization controls. Hybridization, washing and detection were performed using an Affymetrix fluid station. Staining for detection was performed using a Streptavidin-phycoerythrin solution and scanned using an Affymetrix GeneChip scanner following manufacturer's instructions.

6. Identification of Biomarker Genes

The expression values of a total of 54,613 probe sets were available in the CEL files generated by measurements of expression levels obtained from Affymetrix gene chips. Removing constant probe sets and the probesets for which the 95%-quantile does not exceed 7 in log2-scale, 19,078 probesets remained (i.e., probesests for which at least 5% of the sample values are above 7.). The remaining probe sets were subsequently mapped to gene symbols. For probeset on the gene chip, it is associated with one gene symbol but for a given gene symbol, several probesets may correspond to it. To choose a representative probeset for a given gene symbol, the interaction term and its p-value is calculated. Among the probesets corresponding to a given gene symbol, the one with the lowest p-value is used.. After this process, 10'966 genes remained and were used for further analysis. The estimation for each gene was performed using the R-package *limma*. As can be seen in Figure 3, many genes show a significant differential effect due both to the overall smoking effect and tissue difference.

An interaction model was used to estimate the effect of smoke exposure and tissue response as well as the effects of their interactions. An interaction effect is a change in the simple main effect of one variable over levels of the second variable. In this analysis, positive interaction effect means that the effect of smoke exposure in lung is larger in tumor tissue than in healthy tissue. The objective here is to detect a change between tumor and healthy tissues that are different in smokers as compared to the change between tumor and healthy tissues in non-smokers. So the linear regression model used for each gene is :

Log2expression value = b0+b1*SmokingStatus+b2*Tissue+b3 * Tissue:SmokingStatus+ epsilon.

The gene expression values (log2-scale) are modeled as an intercept (b0) plus a smoking status effect (b1) plus a tissue effect (b2) plus an interaction term effect plus a residual term (b3).

The regression coefficients are estimated by the method of ordinary least squares which minimizes the sum of squared vertical distances between the observed responses in the dataset and the responses predicted by the linear approximation.
i.e., minimizing ||gene -(b0+b1*SmokingStatus+b2*Tissue+b3 * Tissue:SmokingStatus) ||^2

For each coefficient, it is tested if it is non-zero H0:bi=0. It is computed under H0 bi/sd(bi) which will follow a t-distribution (Student). The significance is assessed by moderated t-statistics [Smyth, G. K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Statistical Applications in Genetics and Molecular Biology 3, No. 1, Article 3]. In all cases, we used as cutoff values a log ratio > 0.35 and p <0.05.

The results of the statistical analyses identified a total of 111 genes (including the three high priority markers identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1)) that are specifically upregulated (88 genes) or down regulated (23 genes) in lung adenocarcinomas from smokers when compared to the lung adenocarcinoma of non-smokers.

7. Prediction of smoking status on an independent dataset.

Forward linear discriminant analysis (FLDA) was applied to the dataset in order to identify a sparse gene signature and build a classifier, which was then utilized to predict the smoking status of samples in an independent data set from Landi et al. (GSE10072).

The FLDA takes as input the_molecular profile, typically gene expression levels, as well as the phenotype of interest. In our study the phenotype of interest is the smoking status. Given a gene expression profile matrix X with m samples and n genes and the phenotype vector, a predictive signature is derived from the following steps. Firstly, a moderated t-test for the two phenotypes (smoking and non-smoking) is performed for each gene (using the R-package *limma)*. The genes are reordered

according to the decreasing order of the absolute value of the statistics t; (2) Secondly, for all j, the j genes with the highest t-statistic, in absolute value, are selected for training a classifier. Here, a Linear Discriminant Analysis (LDA) was used. The average of the Matthews correlation coefficient (MCC) obtained in a cross validation experiment (five times, five-fold) is recorded. The process above is repeated for all j's. Lastly, the index for which the average MCC is maximum will serve to select the genes that will constitute the gene signature The final prediction model is the trained on this set of genes with LDA.

Formula for the calculation of the MCC value:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FN)(TN + FP)(TP + FP)(TN + FN)}}.$$

Despite the fact that both datasets were generated in different Affymetrix platforms (Human Genome U133 Plus 2.0 Array in the present study vs Human Genome U133A Array in the Landi dataset), which makes it harder to predict the smoking status of samples, FLDA identified a gene signature composed of three genes: FEN1, RIPK2, and CCNE1 which are strongly involved in the regulation of cell proliferation and survival. An LDA model is trained on data of the present study. The heatmap of this gene signature in the training data is shown in Figure 8. By applying this model, making sure that the means for each gene are equal (centering both datasets), the smoking status for samples in the independent data set was predicted. The sensitivity, specificity, and MCC are 0.63, 0.75, 0.37 respectively.

8. Biomarker genes:

Genes involved in modulation of the immune response in terms of immune evasion of the lung adenocarcinoma

Compared to lung adenocarcinoma of non-smokers, lung adenocarcinoma in smokers is characterized by the up-regulation of genes that are involved in cell-to-cell signaling and the modulation of the immune response: STAT1, IFNAR2, PSME2, PSMB9, CXCL9, CXCL10, CD274, IL15, IL15RA, IL6, CCL5, CCL4, IL18RAP, IL2RB (Figure 5). The overall effect of this group of genes is to promote immune evasion of the lung carcinoma.

Genes involved in apoptosis and cell survival network

Compared to lung adenocarcinoma of non-smokers, lung adenocarcinoma in smokers are characterized by the up-regulation of CDK1, FEN1, CCNE1, MCM4, CDC20, BIRC5, CDK4 and GADD45B. The overall effect of these genes is to promote tumor growth by up-regulating processes involved in cell proliferation: DNA replication, cell cycle progression, mitosis, as well the repair of DNA damage (Figure 6).

Genes involved in cellular growth and proliferation network

Finally, compared to lung adenocarcinoma of non-smokers, lung adenocarcinoma from smokers are characterized by the up-regulation of: P2RY6, CFLAR, BIRC3, NFKB2, CHEK1, CXCL12, SPP1, LILRB2, ADA, RIPK2 and the down-regulation of SORT1. (figure 2). The overall effect of the activity of this network is to promote tumor cell survival and protect tumor cells from the activation of apoptosis and other cell death pathways.

Genes reportedly having a tumor suppressor effect when over-expressed in lung and other cancers

It was surprising to find, that some of the genes resulting from the statistical analysis were reported in the literature as having a role in tumor suppression when over-expressed in lung or other cancer tissues. For example:

PSMB9: Different studies have demonstrated a deficiency / lack of expression of MHC class I molecules in the surface of many types of tumors, (Singal et al., 1996, Delp et al., 2000). PSME9 has been postulated as a tumor suppressor in cancer of the uterus (Hayashi et al., 2011)

PSME2: Similar to PSMB9, lower levels are associated with increased aggressiveness and resistance to therapy in melanoma (Stone et al., 2009) and gastric adenocarcinomas (Zheng et al., 2012).

CCL4: Increased levels of CCL4 (MIP1b) are reported to decrease the number of metastasis in murine cell lymphoma (Menten et al., 2002) and a mouse model of lung cancer (Ishihara et al., 1998).

CCL5: Increased levels of CCL5 (RANTES) are associated with better prognosis and increased survival in lung adenocarcinoma (Moran et al., 2002, Ohri et al., 2010)

IFNAR2:. Increased levels of IFNAR2 are associated with tumor suppression (Mendoza-Villanueva et al., 2008, Swann et al., 2007) yet IFNAR2 overexpression

was observed in various histological types of lung cancer, and appears to be associated with lung cancers that behave aggressively (Tanaka et al., 2012). This is an example of a gene that apparently plays different roles in different settings.

CXCL9: Increased levels of CXCL9 (MIG) are reported to inhibit cell growth in mouse and human lung tumors (Addison et al., 2000, Winter et al., 2007).

CXCL10: Treatment with CXCL10 (INP-10) is reported to inhibit metastasis and to increase cell death in a mouse model of lung tumor (Arenberg et al., 2001) and in human non-small cell lung cancer (Ohri et al., 2010).

IL18RAP: No conclusive data are provided in the literature on the role of for IL18 receptor.

IL2RB: Decreased levels of IL2RB are reported to increase the number of metastasis in a mouse model of malignant pleural effusion using human adenocarcinoma cells (Arenberg et al., 2001).

IL15: Tumor nodule formation was retarded and tumor growth was inhibited after treatment with IL15 (Tang et al., 2008).

STAT1: Activation of STAT1 pathway is reported to inhibit lung cancer cell proliferation, anchorage-independent growth, tumorigenesis, cell motility, and invasion, and to slow cell cycle progression (Tsai et al., 2006). Stat1 is further reported to inhibit lung tumor formation by activated K-Ras (Wang et al., 2008).

Table 1: List of biomarker genes specifically upregulated in lung adenocarcinomas of smokers when compared to non-smokers

Gene symbol	Gene name	probeset_id	Log fold change
IFNAR2	interferon (alpha, beta and omega) receptor 2	204785_x_at	0.40900
SNRPG	small nuclear ribonucleoprotein polypeptide G	205644_s_at	0.44000
PSMB7	proteasome (prosome, macropain) subunit, beta type, 7	200786_at	0.46300
POLR3F	polymerase (RNA) III (DNA directed) polypeptide F, 39 kDa	205218_at	0.48300
MARS	methionyl-tRNA synthetase	213671_s_at	0.49800
TGS1	trimethylguanosine synthase 1	238346_s_at	0.49800
GTF2E1	general transcription factor IIE, polypeptide 1, alpha 56kDa	205930_at	0.50300

PSMD12	proteasome (prosome, macropain) 26S subunit, non-ATPase, 12	202352_s_at	0.50900
POLR2H	polymerase (RNA) II (DNA directed) polypeptide H	209302_at	0.53800
POLR2E	polymerase (RNA) II (DNA directed) polypeptide E, 25kDa	217854_s_at	0.55100
TAF5	TAF5 RNA polymerase II, TATA box binding protein (TBP)-associated factor, 100kDa	210053_at	0.56100
CDK4	cyclin-dependent kinase 4	202246_s_at	0.56900
SNRPD1	small nuclear ribonucleoprotein D1 polypeptide 16kDa	202690_s_at	0.58700
TARS	threonyl-tRNA synthetase	201263_at	0.59900
GTF2E2	general transcription factor IIE, polypeptide 2, beta 34kDa	202680_at	0.60200
EIF4A3	eukaryotic translation initiation factor 4A3	201303_at	0.63300
GART	phosphoribosylglycinamide formyltransferase, phosphoribosylglycinamide synthetase, phosphoribosylaminoimidazole synthetase	212378_at	0.65100
PSME2	proteasome (prosome, macropain) activator subunit 2 (PA28 beta)	201762_s_at	0.66300
PVRL2	poliovirus receptor-related 2 (herpesvirus entry mediator B)	203149_at	0.67700
SNRPF	small nuclear ribonucleoprotein polypeptide F	203832_at	0.68900
MED10	mediator complex subunit 10	223247_at	0.69600
NBN	Nibrin	202907_s_at	0.70600
TCEB1	transcription elongation factor B (SIII), polypeptide 1 (15kDa, elongin C)	202824_s_at	0.70600
TUBB3	tubulin, beta 3 class III	202154_x_at	0.70900
PPP3CB	protein phosphatase 3, catalytic	202432_at	0.71300

	subunit, beta isozyme		
CFLAR	CASP8 and FADD-like apoptosis regulator	210563_x_at	0.71400
MCM6	minichromosome maintenance complex component 6	201930_at	0.72300
LYN	v-yes-1 Yamaguchi sarcoma viral related oncogene homolog	202625_at	0.72900
ITGA5	integrin, alpha 5 (fibronectin receptor, alpha polypeptide)	201389_at	0.79100
CDC34	cell division cycle 34 homolog (S. cerevisiae)	212540_at	0.79900
FYN	FYN oncogene related to SRC, FGR, YES	210105_s_at	0.88800
SNRPB	small nuclear ribonucleoprotein polypeptides B and B1	213175_s_at	0.90500
CXCL12	chemokine (C-X-C motif) ligand 12	203666_at	0.91200
E2F5	E2F transcription factor 5, p130-binding	221586_s_at	0.92600
RFC3	replication factor C (activator 1) 3, 38kDa	204127_at	0.93300
NFATC1	nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 1	210805_s_at	0.93800
LCP2	lymphocyte cytosolic protein 2 (SH2 domain containing leukocyte protein of 76kDa)	205269_at	0.94100
RIPK2	receptor-interacting serine-threonine kinase 2	209545_s_at	0.95600
RPL22L1	ribosomal protein L22-like 1	225541_at	0.99800
P2RY6	pyrimidinergic receptor P2Y, G-protein coupled, 6	208373_s_at	1.00600
ADA	adenosine deaminase	204639_at	1.03800
PAICS	phosphoribosylaminoimidazole carboxylase, phosphoribosylaminoimidazole succinocarboxamide synthetase	201013_s_at	1.04000

STAT1	signal transducer and activator of transcription 1, 91kDa	200887_s_at	1.05600
IL18RAP	interleukin 18 receptor accessory protein	207072_at	1.05800
PTTG1	pituitary tumor-transforming 1	203554_x_at	1.06400
ODC1	ornithine decarboxylase 1	200790_at	1.09700
IL15RA	interleukin 15 receptor, alpha	207375_s_at	1.09800
NFKB2	nuclear factor of kappa light polypeptide gene enhancer in B-cells 2 (p49/p100)	207535_s_at	1.13000
WARS	tryptophanyl-tRNA synthetase	200629_at	1.13800
OSMR	oncostatin M receptor	226621_at	1.16900
TAP2	transporter 2, ATP-binding cassette, sub-family B (MDR/TAP)	225973_at	1.21000
CD3D	CD3d molecule, delta (CD3-TCR complex)	213539_at	1.23400
TAP1	transporter 1, ATP-binding cassette, sub-family B (MDR/TAP)	202307_s_at	1.23400
SLC7A5	solute carrier family 7 (amino acid transporter light chain, L system), member 5	201195_s_at	1.25200
IFITM1	interferon induced transmembrane protein 1	214022_s_at	1.25700
IL2RB	interleukin 2 receptor, beta	205291_at	1.26600
S1PR3	sphingosine-1-phosphate receptor 3	228176_at	1.29000
FEN1	flap structure-specific endonuclease 1	204767_s_at	1.34100
CCNE1	cyclin E1	213523_at	1.37000
GADD45B	growth arrest and DNA-damage-inducible, beta	207574_s_at	1.37900
FLT1	fms-related tyrosine kinase 1 (vascular endothelial growth factor/vascular permeability factor receptor)	222033_s_at	1.40300
CDK1	cyclin-dependent kinase 1	203213_at	1.45400

BUB1	budding uninhibited by benzimidazoles 1 homolog (yeast)	209642_at	1.49500
SPP1	secreted phosphoprotein 1	209875_s_at	1.51200
BIRC3	baculoviral IAP repeat containing 3	210538_s_at	1.51400
PSMB9	proteasome (prosome, macropain) subunit, beta type, 9 (large multifunctional peptidase 2)	204279_at	1.53500
MCM4	minichromosome maintenance complex component 4	222036_s_at	1.54100
VCAM1	vascular cell adhesion molecule 1	203868_s_at	1.55200
CHEK1	checkpoint kinase 1	205394_at	1.59500
PMAIP1	phorbol-12-myristate-13-acetate-induced protein 1	204285_s_at	1.60100
TYMS	thymidylate synthetase	202589_at	1.62700
TNFAIP3	tumor necrosis factor, alpha-induced protein 3	202644_s_at	1.63900
IL15	interleukin 15	217371_s_at	1.65500
UBE2C	ubiquitin-conjugating enzyme E2C	202954_at	1.76500
LILRB2	leukocyte immunoglobulin-like receptor, subfamily B (with TM and ITIM domains), member 2	207697_x_at	1.79500
CCL8	chemokine (C-C motif) ligand 8	214038_at	1.83900
CCL4	chemokine (C-C motif) ligand 4	204103_at	1.84200
CDC20	cell division cycle 20 homolog (S. cerevisiae)	202870_s_at	1.85500
ITK	IL2-inducible T-cell kinase	211339_s_at	1.87200
CD274	CD274 molecule	227458_at	1.89900
CCL5	chemokine (C-C motif) ligand 5	204655_at	1.94100
BIRC5	baculoviral IAP repeat containing 5	202095_s_at	1.96000
AIM2	absent in melanoma 2	206513_at	2.08100
CXCL10	chemokine (C-X-C motif) ligand 10	204533_at	3.08400

CXCL9	chemokine (C-X-C motif) ligand 9	203915_at	3.18500
CXCL11	chemokine (C-X-C motif) ligand 11	210163_at	3.20100
IL6	interleukin 6 (interferon, beta 2)	205207_at	3.33600
GZMB	granzyme B (granzyme 2, cytotoxic T-lymphocyte-associated serine esterase 1)	210164_at	3.53000

Table 2: List of biomarker genes specifically downregulated in lung adenocarcinomas of smokers when compared to non-smokers

Gene symbol	Gene name	probeset_id	Log fold change
ALDH6A1	aldehyde dehydrogenase 6 family, member A1	221588_x_at	-1.318
ABAT	4-aminobutyrate aminotransferase	209460_at	-1.281
ACSL5	acyl-CoA synthetase long-chain family member 5	222592_s_at	-1.207
FUCA1	fucosidase, alpha-L- 1, tissue	202838_at	-1.089
GALC	Galactosylceramidase	204417_at	-0.943
ALDH3A2	aldehyde dehydrogenase 3 family, member A2	202053_s_at	-0.912
GUSB	glucuronidase, beta	202605_at	-0.778
SORT1	sortilin 1	224818_at	-0.776
MANBA	mannosidase, beta A, lysosomal	203778_at	-0.744
HSD17B4	hydroxysteroid (17-beta) dehydrogenase 4	201413_at	-0.734
IGF2R	insulin-like growth factor 2 receptor	201393_s_at	-0.711
NAGLU	N-acetylglucosaminidase, alpha	204360_s_at	-0.704
ATP6V0A1	ATPase, H+ transporting, lysosomal V0 subunit a1	212383_at	-0.699
CAT	Catalase	201432_at	-0.677
HEXB	hexosaminidase B (beta polypeptide)	201944_at	-0.635
SCP2	sterol carrier protein 2	211733_x_at	-0.620
HEXA	hexosaminidase A (alpha polypeptide)	201765_s_at	-0.592
ATP6V0D1	ATPase, H+ transporting, lysosomal 38kDa, V0 subunit d1	212041_at	-0.573
ALDH9A1	aldehyde dehydrogenase 9 family, member A1	201612_at	-0.557
ATP6AP1	ATPase, H+ transporting, lysosomal accessory protein 1	207809_s_at	-0.543
ASAH1	N-acylsphingosine amidohydrolase (acid ceramidase) 1	213702_x_at	-0.514
CPT2	carnitine palmitoyltransferase 2	204264_at	-0.497
GNS	glucosamine (N-acetyl)-6-sulfatase	212334_at	-0.460

Table 3: List of biomarker genes specifically upregulated in lung adenocarcinomas from smokers when compared to non-smokers and reported in the prior art to have a tumor suppressor effect when overexpressed in lung and other cancers.

STAT1	signal transducer and activator of transcription 1, 91kDa	200887_s_at	1.05600
PSME2	proteasome (prosome, macropain) activator subunit 2 (PA28 beta)	201762_s_at	0.66300
PSMB9	proteasome (prosome, macropain) subunit, beta type, 9 (large multifunctional peptidase 2)	204279_at	1.53500
CXCL9	chemokine (C-X-C motif) ligand 9	203915_at	3.18500
CXCL10	chemokine (C-X-C motif) ligand 10	204533_at	3.08400
IL15	interleukin 15	217371_s_at	1.65500
CCL5	chemokine (C-C motif) ligand 5	204655_at	1.94100
CCL4	chemokine (C-C motif) ligand 4	204103_at	1.84200
IL18RAP	interleukin 18 receptor accessory protein	207072_at	1.05800
IL2RB	interleukin 2 receptor, beta	205291_at	1.26600

9. Identification of Signalling Pathways and Biological Networks

Following a classical approach to gene expression analysis, genes for which the False Discovery Rate (FDR) of the interaction term is below 0.05 are reviewed first. A total of 698 differentially expressed genes were found (coefficient threshold log2(1.2) and FDR threshold 0.05). In order to refine the differentially expressed gene list, a gene set enrichment analysis of the interaction term is performed, which is advantageous as it does not depend on a cutoff value for genes and is able to capture more subtle signals (Mootha et al., Subramanian et al.). Gene set enrichment analysis (GSEA) [11] was used for comparison of the expression of groups of genes (gene-sets) between different tissues. GSEA provides common gene-set patterns even when single gene analysis shows only a few overlapping genes between groups. The differences in gene expression between healthy and tumor tissue in smokers and non-smokers were compared. By doing this double comparison it was possible to investigate the contribution of cigarette smoke to tumor development and progression. Using a stringent FDR cutoff of 0.01, gene-sets were extracted for subsequent analysis. The interpretation of the result was then performed using IPA

(Ingenuity Systems, www.ingenuity.com) and the gene-sets obtained from the intersection between leading edge genes and differentially expressed genes to determine the most significant biological functions and molecular pathways that are specifically activated in smoke-related adenocarcinomas. A differentially expressed gene list containing gene identifiers and corresponding fold changes was first uploaded as an Excel spreadsheet into the IPA software. Each gene identifier was mapped to its corresponding gene object in the Ingenuity Pathways Knowledge Base. These genes were then used as the starting point for pathway analysis. Canonical pathways analysis identified the pathways from the Ingenuity Pathways Analysis library of canonical pathways that were most significant to the data set. The significance of the association between the dataset and the canonical pathway was measured in two ways: (1) ratio of the number of genes from the data set that map to the pathway divided by the total number of genes that map to the canonical pathway was displayed, (2) Fischer's exact test was used to calculate a p-value determining the probability that the association between the genes in the dataset and the canonical pathway is explained by chance alone. Only molecules from the dataset that met the cut-off criteria (1.5 fold-change and $p<0.05$) were considered for the analysis.

Figure 4A shows the top ten biological functions associated with lung tumors of smokers. As expected, these include functions related to cancer, but also to cellular growth and proliferation, cell death and survival, as well as different aspects of the immune response, such as immune cell trafficking, inflammatory response, and cellular movement. Moreover, the gene-sets were categorized into canonical pathways. Figure 4B shows the top ten canonical pathways activated specifically in CS-related lung adenocarcinomas. Consistent with the previous figure, the most significant pathways are associated with the DNA damage response and the regulation of cell cycle progression, proliferation and apoptosis.

There is growing evidence indicating that the molecular pathways driving tumorigenesis in smokers and non-smokers are quite different, the former generally being more aggressive, associated with poor survival and resistance to therapy. Accordingly, it was found that lung tumors in smokers show increased activation of cell proliferation and survival pathways, as well as alterations in the immune response. To gain further insight into the molecular mechanisms driving CS-related

tumor growth, the gene-sets were used to generate *de novo* networks based on knowledge from the literature contained in the IPA knowledge base.

Lung adenocarcinoma is a heterogeneous disease, with specific clinical and genetic features depending on the smoking-status of the patients. Several studies indicate that the currently identified genetic mutations cannot explain the full spectrum of phenotypes observed in the clinics, suggesting a broader and more complex network of interaction among the various components. In the present study, the focus was on biological processes, rather than on individual molecules, which drive lung tumorigenesis in smokers. In particular, it was found that there is a strong activation of immune regulatory pathways in CS-induced tumors. Numerous studies demonstrated antitumoral role for both the innate and the adaptive immune responses, thus supporting the concept of immunosurveillance (Schroder et al, Vesely et al.). Further studies, however, revealed a more complex scenario where the immune system exerts a selective pressure over the nascent tumor cells, which shapes or edits tumor outgrowth, a more global concept known as immunoediting (Schroder et al, Vesely et al.). When the immune system fails to detect and / or control transformed cells, the tumor is able to survive and continue growing. While in some cases this is the result of compromised immune response (i.e. immunosuppression after an organ transplant), more and more evidence supports a more active role of the tumor in the immune evasion process, either by inhibiting tumor recognition and cytolysis by immune effector cells, by inducing immunotolerance, or by creating an immunosuppressive tumor microenvironment (Zaidi et al., Schroder et al., Manjili et al.). Our results suggest that multiple immune response pathways are upregulated in smoke-related lung tumors. Individually, most of these cell signaling pathways would prevent tumor growth, however, unexpectedly their combined actions, together with the effects of CS exposure results in an intricate immune response that allows immune evasion and favors tumor cell proliferation. Tumor growth is further enhanced by overactivation of cell survival pathways and inhibition of apoptosis. The findings are of potential clinical interest. On one hand, they can provide the basis for the identification of novel therapeutic targets in the treatment of lung adenocarcinoma. On the other hand, they could help select the most appropriate treatment for the patient, thus maximizing the chances for success to therapy.

Cell-to-cell signaling network

The first network contains a number of molecules that are essential in cell-to-cell signaling and the modulation of the immune response (Figure 5). The central molecules are: IL15 (interleukin 15), IL6 (interleukin 6) and STAT1 (signal transducer activator of transcription 1). IL15 is a pro-inflammatory cytokine expressed in many tissues including lung epithelium. Some of its downstream signaling elements include JAK/STAT, MAPK and PI3K/AKT pathways, which promote the proliferation and activation of natural killer (NK) cells as well as T and B lymphocytes (Carson et al (1994 and 1997), Schluns et al.) thus initiating a strong immunological response. This response is further enhanced by STAT1, an essential mediator of interferon gamma (IFNG), and its downstream targets IL6, CCL4 (C-C motif ligand 4) and CCL5 (C-C motif ligand 5). These cytokines and chemokines further promote the recruitment of lymphocytes, monocytes and NK cells. IL15 and STAT1 also play an important role in the development antitumoral immune responses. In fact, due to its ability to stimulate strong T-cell mediated cytotoxic responses IL15 is currently being evaluated as an immunotherapeutic agent (Le Maux Chansac et al., Takeuchi et al., Teague et al.). Furthermore, STAT1 can induce growth arrest and apoptosis in cancer cells through the activation of p27 (Wang et al.), $p21^{WAF}$ and caspases (Yu et al.).

Because of their tumor suppressive effects, immune mediators are often inhibited or attenuated in lung and other malignancies. Indeed, this is the case in tumors from non-smokers, where downregulation was observed in the expression of *IL15*, *STAT1*, *CCL4* and *CCL5* genes in tumors compared to healthy tissue. Surprisingly, and despite generally being more aggressive, lung tumors in smokers showed an upregulation of these four genes when compared to non-smokers.

In order to better understand the role of IL15 and STAT1 in the development and progression of lung adenocarcinomas in smokers, the network was evaluated as a whole, taking into account multiple interactions among the different nodes. The rationale for this approach is that immune regulators show a high level pleiotropy and crosstalk among the different signaling pathways which may lead to different outcomes depending on the cellular context. IL15, for example, can be found soluble, but the majority of the protein is bound to the cellular membrane, either directly or through an interaction with IL15RA (IL15 receptor subunit alpha) (Jakobisiak et al.). Pathway activation requires the binding of IL15 to the IL15 receptor complex. Under normal circumstances, this complex is formed by the alpha subunit in the host cell and IL2RB and IL2RG (IL2 receptor subunits beta and gamma, respectively) in the

target cell. In renal and other carcinomas, however, IL15 can bind to an aberrant form of the receptor complex formed by alpha and beta chains alone. This aberrant conformation then promotes epithelial to mesenchymal transition (Giron-Michel et al.) and cell proliferation and transformation (Motegi et al.). Formation of this aberrant receptor is favored by increased levels of IL15 and the alpha and beta subunits of its receptor (Giron-Michel et al., Khawam et al.). Although increased gene expression does not necessarily translate into increased protein levels, the observed upregulation of IL15, IL15RA and IL2RB genes in lung tumors from smokers strongly suggests the presence of an aberrant IL15 signaling in lung tumors from smokers.

IL15 also increases the expression of IL18RAP (interleukin 18 receptor accessory protein), an enhancer of IL18 receptor signaling that promotes NFKB activation (Born et al.). IL15/IL18RAP activation induces the release of IFNG (Sareneva et al.), which, in turn, regulates the expression of many genes either directly or through activation of JAK/STAT pathways. Interferons are a class of cytokines that modulate innate and adaptive immune responses against viruses, bacteria and tumor cells through an effect on NK cells and cytotoxic T lymphocytes. In particular, INFG has important immunomodulatory functions, including T-cell differentiation, activation and homeostasis, NK cell activation, lysosomal activation in macrophages, and the promotion of antigen presentation by upregulating the expression of components of the class I major histocompatibility complex (MHC) and of the immunoproteasome (Schroder et al.). A priori, all these effects support a role for IFNG in tumor cell identification and elimination. Indeed, a number of studies suggest that IFNG plays a key role in tumor surveillance by increasing tumor immunogenicity (Dunn et al., Zaidi et al.), inhibiting tumor cell proliferation and promoting apoptosis (Zaidi et al., Ikeda et al.). In fact, the therapeutic use of IFNG has already been tested in the clinic with melanoma patients, although it showed low efficacy (Schiller et al.) and surprisingly, even faster disease progression compared to non-treated patients (Meyskens et al.). In the recent years, a number of studies have demonstrated a dual role for IFNG and the immune response on tumor development and progression. While IFNG has the potential to drive an antitumor immune response, increased / sustained IFNG activation is able to inhibit apoptosis and promote tumor cell proliferation and metastasis (Zaidi et al.).

In the present study, it was not possible to detect changes in IFNG gene expression *per se*. However, by using *INTERFEROME* database (Samarajiwa et al.), it was

possible to identify a high number of IFNG target genes upregulated in lung tumors from smokers compared to non-smokers (Figure 5), which strongly supports the idea of a sustained IFNG activation. Increased expression of CCL4, CCL5 and the C-X-C motif chemokines CXCL9 and CXCL10 in tumors from smokers were observed. These molecules are potent regulators of monocytes / macrophages, T-lymphocytes, dendritic cells and NK cell activation. IFNG-mediated overexpression of these proteins is associated with alveolar destruction and emphysema (Ma et al.). Elevated levels of CXCL9 and CXCL10 can also be detected at metastatic sites in lung and colorectal carcinomas, where they promote MMP9 (matrix metalloprotease 9) mediated disruption of the endothelial barrier and cellular migration. Interestingly, CS dramatically upregulates IFNG (Ma et al.) and thus, the levels of CXCL9 and CXCL10. Increased levels of these chemokines will likely result in more tissue destruction and putatively a higher risk for invasion and metastasis.

Previous reports have shown that IFNG allows tumors to escape to immune surveilance.by upregulating the expression of MHC-I and MHC-II molecules, leading to decreased NK-mediated celly lysis and T-cell mediated apoptosis (Zaidi et al., Maio et al., Hemon et al, Beatty et al.). Similarly, in lung tumors from smokers an upregulation was observed of several genes involved in antigen presentation by MHC-I and that are downstream targets of IFNG (Figure 5). It was hypothesized that these tumors have active mechanisms that allow them to escape elimination by the immune system. IFNG upregulates the expression of IFNAR2 (interferon alpha/beta receptor beta chain) leading to activation of JAK/STAT1 pathways and IL6 overexpression. IFNAR2 also promotes the activation of the immunoproteasome, thus increasing antigen presentation in tumor cells. The immunoproteasome is further activated by the upregulation of two other relevant genes, PSME2 (proteasome activator complex subunit 2) and PSMB9 (proteasome subunit beta type-9). Thus, overactivation of the immunoproteasome and excessive antigen presentation can, in fact, induce a downregulation of the immune response. NK cells respond to infected or transformed cells either by killing the abnormal cells or by releasing immunomodulatory chemokines and cytokines such as IFNG. In order to prevent damage to normal tissues, NK cells are normally restrained by different types of inhibitory receptors that recognize MHC class I molecules in the target-cell. When MHC-I expression is low, NK cells are liberated from the inhibitory receptors and can kill target cells more efficiently (French et al.). Conversely, high expression of PSME2

and PSMB9 leads to increased MHC-I expression and increases in the chances of interaction with inhibitory receptors in the NK cell, thus preventing tumor cell elimination. Therefore, upregulation of MHC-I molecules would be a very ingenious mechanism for tumors to evade immune surveillance. This type of response has been already reported in hematopoietic malignancies and is mediated by STAT1 (Kovacic et al.). Furthermore, CD274 was found (also known as programmed cell death ligand 1, PD-L1) also upregulated in our network. In response to IFNG, CD274 sends an inhibitory signal that inhibits the proliferation of activated T-lymphocytes, monocytes NK and dendritic cells (Riley et al.). The main role of CD274 is to induce fetomaternal tolerance during pregnancy. CD274 overexpression correlates with poor prognosis in non-small cell lung cancer (NSCLC), where induces immune escape by preventing the maturation of dendritic cells (Mu et al.). Finally, high levels of IFNAR2, PMSE2 and PMSB9 lead to sustained activation of p38/MAPK and NFKB pathways. As a central mediator of inflammation, NFKB can promote cellular death or survival. NFKB regulates cellular fate in connivance with the tumor suppressor gene TP53. TP53 mutations can be found in lung adenocarcinomas from both smokers and non-smokers, but the percentage is significantly higher in smokers, and can be directly linked to CS exposure (Couraud et al.). While in a p53 proficient cell, sustained NFKB would cause growth arrest and apoptosis, in lung tumors from smokers the opposite effect (increased proliferation and cell survival) would be expected.

Cell proliferation and survival network

A number of upregulated of genes was found that control cell proliferation and survival in CS-related lung tumors. Tumors are by definition, an abnormal and uncontrolled growth of a specific tissue. Therefore it is not surprising to observe activation of molecular pathways that promote cell proliferation and /or prevent cellular death. The interesting observation, however, is that these genes are upregulated not only in tumor compared to healthy tissue, but more importantly in lung tumors from smokers compared to non-smokers. In addition, most of the upregulated genes have been previously associated with poor outcome in multiple types of cancer, including those of the lung, supporting the more aggressive phenotype often observed in CS-related lung tumors.

IPA was used to build a cell proliferation and survival network and to investigate the interactions among its components. The network is divided into two subnetworks. At

the core of the first subnetwork (Figure 6A) lies CDK1, (cyclin dependent kinase 1) the only mitotic kinase that is both necessary and sufficient to drive cell division (Santamaria et al., Linares et al.). CDK1 interacts directly with other key regulators of cell cycle progression, such as *CDC20* (cell divison cycle 20 homolog), *BIR5* (baculoviral inhibitor of apoptosis repeat-containing 5 or survivin) and *FEN1* (Flap structure-specific endonuclease 1). CDC20, a member of the anaphase promoting complex, is responsible for the inhibition / degradation of many cell cycle regulators like $p21^{ARF}$ (i.e. inhibits CDK1 complex formation, cyclin E, Cyclin D and PCNA) (Kato et al., Qiao). *BIRC5* is an inhibitor of apoptosis by inhibiting BAX and FAS-mediated pathways (Tamm et al.) and a transcriptional inhibitor of the P21WAF1 gene (Tang et al.). FEN1 is involved in the processing of the Okazaki fragments in the DNA lagging strand synthesis (Henneke et al.), telomere stability (Saharia et al.) and DNA-repair pathways (Klungland et al.), where is considered a limiting factor. Another important component of the network is CCNE1 (Cyclin E1) which serves as a regulator of cyclin-dependent kinases and the RB family of proteins, thus promoting G1/S transition and cell cycle progression (Harbour et al., Shanahan et al.). CCNE1 overexpression is particularly common in smoke-related lung tumors where it shortens cell cycle and promotes genomic instability (Ohtsubo et al., Spruck et al.). The presence of MCM4 (minichromosome maintenance protein 4) is also not surprising in this network, as it is considered a marker of poor prognosis in NSCLC and other tumors (Freeman et al., Kikuchi et al.). As a key component of the pre-replication complex, MCM4 is essential for genome replication (You et al.). Uncontrolled proliferation is further promoted by overexpression of CDK4 (cyclin dependent kinase 4). CDK4 promotes G1/S transition, partially by phosphorylating RB1, which may explain at least in part, the impairment of the retinoblastoma (RB) pathway known to occur in most lung tumors (Wikman et al.). Interestingly, an important inhibitor of this pathway, CDKN2A, is most often mutated in lung adenocarcinomas (Ding et al., Imielinski et al.), suggesting the existence of constitutive activated cyclin-dependent growth signals.

The effect of this network is to boost tumor cell proliferation. Exaggerated cell growth increases the risk for accumulating DNA mutations leading to genomic instability and compromised cellular function. In fact, many tumors show both high rates of cell proliferation and apoptosis, and the growth rate of the tumor depends on the ratio between these two processes. This is not the case for CS-related lung tumors, where

higher proliferative potential is accompanied by activation of different pathways promoting cell survival, as depicted in the second subnetwork (Figure 6B).

On one hand, the upregulation of genes with a key role in DNA repair was observed, such as FEN1 and CHEK1. Besides its role in DNA replication, FEN1 participates in the repair of different types of DNA damage, including replication-induced single and double-strand DNA breaks and it is considering a limiting factor in DNA repair (Nikolova et al.). CHEK1 is a critical component of DNA replication, intra-S phase, G2/M transition, and mitotic spindle-assembly checkpoints (Bartek et al.). In response to DNA damage, it becomes activated and blocks cell cycle progression until the damage is repaired (Bartek et al., Peng et al.). If the DNA lesion cannot be finally repaired, normal cells activate senescence/cell death programs to prevent further damage. However, when tumor suppressor genes such as TP53 and RB family members are mutated, as it is often the case in lung and other types of cancer, the activation of these programs is compromised and the cells are more likely to escape cell death.

On the other hand, it was found that tumors of smokers activate pathways that inhibit apoptosis and promote cell survival (Figure 6B). SPP1 (secreted phosphoprotein 1 or osteopontin) is an important bone remodeling factor often overexpressed in tumors of the lung, breast and colon that is associated with poor prognosis (Shevde et al.). SPP1 is also upregulated in response to CS (Bishop et al.) and has been suggested to promote preneoplastic cell growth *in vivo* and *in vitro* (Pazolli et al.). SPP1 is able to inhibit BAX activation and prevent BCL2 downregulation, thus inhibiting caspase-9 and caspase-3 dependent cell apoptosis (Gu et al.). Based on the present network, cell survival would be further reinforced by the overexpression of the caspase inhibitors CFLAR (CASP8 and FADD-like apoptosis regulator) and BIRC3 (baculoviral inhibitor of apoptosis repeat-containing 3). Furthermore, constitutive activation of NFKB, either through activating mutations or downregulation of its inhibitors (IKBs) is commonly found in lung and other tumors and often associated with increased resistance to therapy (Rayet et al.). NFKB may be upregulated in tumors from smokers. First, *NFKB2* (the p100 subunit of NFKB) is upregulated in smokers and its expression is directly induced by CS. Second, RIPK2 (Receptor-interacting serine/threonine-protein kinase 2) is a potent activator of NFKB through the inhibition of ikbs complexes (Kim et al.). Finally, BIRC3 and BIRC5 have recently been identified as the E3-ubiquitin ligases responsible for of RIPK2 ubiquitinilation, a

prerequisite for NFKB activation (Bertrand et al.). The combined effect of these two subnetworks is to provide the tumor with a formidable proliferative potential, while protecting cancer cells from any attempt of activation of senescence or apoptosis programs that may limit its growth.

In order to investigate whether the regulation of the immune response and cell-to-cell signaling and the cell proliferation and survival networks represent independent processes or they are being regulated in a coordinated manner, upstream regulator analysis in IPA was performed and identified several candidates. From all of them, IFNG, TNF (tumor necrosis factor) and IL1B (interleukin 1 beta) were selected based on their z-score, p value and number of downstream targets to expand our networks and find possible interactions between them. Although IFNG, TNF, or IL1B were not upregulated in tumors of smokers compared to non-smokers (likely due to high level of expression in both tumor tissues), they interact strongly with multiple components of our networks (Figure 7) suggesting not only a major role for all three molecules but also that the different activated pathways are acting in a coordinated manner in order to promote tumor growth.

References

ADDISON, C. L., ARENBERG, D. A., MORRIS, S. B., XUE, Y. Y., BURDICK, M. D., MULLIGAN, M. S., IANNETTONI, M. D. & STRIETER, R. M. 2000. The CXC chemokine, monokine induced by interferon-gamma, inhibits non-small cell lung carcinoma tumor growth and metastasis. *Hum Gene Ther,* 11, 247-61.

ARENBERG, D. A., WHITE, E. S., BURDICK, M. D., STROM, S. R. & STRIETER, R. M. 2001. Improved survival in tumor-bearing SCID mice treated with interferon-gamma-inducible protein 10 (IP-10/CXCL10). *Cancer Immunol Immunother,* 50, 533-8.

BARTEK, J. and J. LUKAS, Chk1 and Chk2 kinases in checkpoint control and cancer. Cancer Cell, 2003. 3(5): p. 421-9.

BEATTY, G.L. and Y. Paterson, IFN-gamma can promote tumor evasion of the immune system in vivo by down-regulating cellular levels of an endogenous tumor antigen. J Immunol, 2000. 165(10): p. 5502-8.

BERTRAND, M.J., et al., cIAP1/2 are direct E3 ligases conjugating diverse types of ubiquitin chains to receptor interacting proteins kinases 1 to 4 (RIP1-4). PLoS One, 2011. 6(9): p. e22356.

BISHOP, E., E.H. THEOPHILUS, and I.M. FEARON, In vitro and clinical studies examining the expression of osteopontin in cigarette smoke-exposed endothelial cells and cigarette smokers. BMC Cardiovasc Disord, 2012. 12(1): p. 75.

BORN, T.L., et al., Cloning of a novel receptor subunit, AcPL, required for interleukin-18 signaling. J Biol Chem, 1998. 273(45): p. 29445-50.

CARSON, W.E., et al., A potential role for interleukin-15 in the regulation of human natural killer cell survival. J Clin Invest, 1997. 99(5): p. 937-43.

CARSON, W.E., et al., *Interleukin (IL) 15 is a novel cytokine that activates human natural killer cells via components of the IL-2 receptor.* J Exp Med, 1994. **180**(4): p. 1395-403.

COURAUD, S., et al., Lung cancer in never smokers--a review. Eur J Cancer, 2012. 48(9): p. 1299-311.

DELP, K., MOMBURG, F., HILMES, C., HUBER, C. & SELIGER, B. 2000. Functional deficiencies of components of the MHC class I antigen pathway in human tumors of epithelial origin. *Bone Marrow Transplant,* 25 Suppl 2, S88-95.

DING, L., et al., Somatic mutations affect key pathways in lung adenocarcinoma. Nature, 2008. 455(7216): p. 1069-75.

DUNN, G.P., L.J. Old, and R.D. Schreiber, The three Es of cancer immunoediting. Annu Rev Immunol, 2004. 22: p. 329-60.

FREEMAN, A., et al., Minichromosome maintenance proteins as biological markers of dysplasia and malignancy. Clin Cancer Res, 1999. 5(8): p. 2121-32.

FRENCH, A.R. and W.M. Yokoyama, Natural killer cells and autoimmunity. Arthritis Res Ther, 2004. 6(1): p. 8-14.

GIRON-MICHEL, J., et al., Interleukin-15 plays a central role in human kidney physiology and cancer through the gammac signaling pathway. PLoS One, 2012. 7(2): p. e31624.

GOVINDAN, R., et al., Genomic landscape of non-small cell lung cancer in smokers and never-smokers. Cell, 2012. 150(6): p. 1121-34.

GU, T., et al., Osteopontin is involved in the development of acquired chemo-resistance of cisplatin in small cell lung cancer. Lung Cancer, 2009. 66(2): p. 176-83.

HARBOUR, J.W., et al., Cdk phosphorylation triggers sequential intramolecular interactions that progressively block Rb functions as cells move through G1. Cell, 1999. 98(6): p. 859-69.

HAYASHI, T., HORIUCHI, A., SANO, K., HIRAOKA, N., KASAI, M., ICHIMURA, T., SUDO, T., TAGAWA, Y., NISHIMURA, R., ISHIKO, O., KANAI, Y., YAEGASHI, N., ABURATANI, H., SHIOZAWA, T. & KONISHI, I. 2011. Potential role of LMP2 as tumor-suppressor defines new targets for uterine leiomyosarcoma therapy. Sci Rep, 1, 180.

HEMON, P., et al., MHC class II engagement by its ligand LAG-3 (CD223) contributes to melanoma resistance to apoptosis. J Immunol, 2011. 186(9): p. 5173-83.

HENNEKE, G., S. Koundrioukoff, and U. Hubscher, Phosphorylation of human Fen1 by cyclin-dependent kinase modulates its role in replication fork regulation. Oncogene, 2003. 22(28): p. 4301-13.

IKEDA, H., L.J. OLD, AND R.D. SCHREIBER, The roles of IFN gamma in protection against tumor development and cancer immunoediting. Cytokine Growth Factor Rev, 2002. 13(2): p. 95-109.

IMIELINSKI, M., et al., Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell, 2012. 150(6): p. 1107-20.

ISHIHARA, Y., IIJIMA, H. & MATSUNAGA, K. 1998. Contribution of cytokines on the suppression of lung metastasis. Biotherapy, 11, 267-75.

JAKOBISIAK, M., J. Golab, and W. Lasek, Interleukin 15 as a promising candidate for tumor immunotherapy. Cytokine Growth Factor Rev, 2011. 22(2): p. 99-108.

KALIKAKI, A., et al., Clinical outcome of patients with non-small cell lung cancer receiving front-line chemotherapy according to EGFR and K-RAS mutation status. Lung Cancer, 2010. 69(1): p. 110-5.

KATO, T., et al., Overexpression of CDC20 predicts poor prognosis in primary non-small cell lung cancer patients. J Surg Oncol, 2012. 106(4): p. 423-30.

KHAWAM, K., et al., Human renal cancer cells express a novel membrane-bound interleukin-15 that induces, in response to the soluble interleukin-15 receptor alpha chain, epithelial-to-mesenchymal transition. Cancer Res, 2009. 69(4): p. 1561-9.

KIKUCHI, J., et al., Minichromosome maintenance (MCM) protein 4 as a marker for proliferation and its clinical and clinicopathological significance in non-small cell lung cancer. Lung Cancer, 2011. 72(2): p. 229-37.

KIM, J.Y., et al., TAK1 is a central mediator of NOD2 signaling in epidermal cells. J Biol Chem, 2008. 283(1): p. 137-44.

KLUNGLAND, A. and T. LINDAHL, Second pathway for completion of human DNA base excision-repair: reconstitution with purified proteins and requirement for DNase IV (FEN1). EMBO J, 1997. 16(11): p. 3341-8.

KOVACIC, B., et al., STAT1 acts as a tumor promoter for leukemia development. Cancer Cell, 2006. 10(1): p. 77-87.

LANDI, M.T., et al., Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. PLoS One, 2008. 3(2): p. e1651.

LE MAUX CHANSAC, B., et al., Potentiation of NK cell-mediated cytotoxicity in human lung adenocarcinoma: role of NKG2D-dependent pathway. Int Immunol, 2008. 20(7): p. 801-10.

LINARES, J.F., et al., Phosphorylation of p62 by cdk1 controls the timely transit of cells through mitosis and tumor cell proliferation. Mol Cell Biol, 2011. 31(1): p. 105-17.

LIU, J., et al., Genome and transcriptome sequencing of lung cancers reveal diverse mutational and splicing events. Genome Res, 2012. 22(12): p. 2315-27.

MA, B., et al., Role of CCR5 in IFN-gamma-induced and cigarette smoke-induced emphysema. J Clin Invest, 2005. 115(12): p. 3460-72.

MAIO, M., et al., Reduction in susceptibility to natural killer cell-mediated lysis of human FO-1 melanoma cells after induction of HLA class I antigen expression by transfection with B2m gene. J Clin Invest, 1991. 88(1): p. 282-9.

MANJILI, M.H., Revisiting cancer immunoediting by understanding cancer immune complexity. J Pathol, 2011. 224(1): p. 5-9.

MASSION, P.P., et al., Smoking-related genomic signatures in non-small cell lung cancer. Am J Respir Crit Care Med, 2008. 178(11): p. 1164-72.

MENDOZA-VILLANUEVA, D., DIAZ-CHAVEZ, J., URIBE-FIGUEROA, L., RANGEL-ESCAREAO, C., HIDALGO-MIRANDA, A., MARCH-MIFSUT, S., JIMENEZ-SANCHEZ, G., LAMBERT, P. & GARIGLIO, P. 2008. Gene expression profile of cervical and skin tissues from human papillomavirus type 16 E6 transgenic mice. *BMC Cancer,* 8, 347.

MENTEN, P., SACCANI, A., DILLEN, C., WUYTS, A., STRUYF, S., PROOST, P., MANTOVANI, A., WANG, J. M. & VAN DAMME, J. 2002. Role of the autocrine chemokines MIP-1alpha and MIP-1beta in the metastatic behavior of murine T cell lymphoma. *J Leukoc Biol,* 72, 780-9.

MEYSKENS, F.L., Jr., et al., Recombinant human interferon gamma: adverse effects in high-risk stage I and II cutaneous malignant melanoma. J Natl Cancer Inst, 1990. 82(12): p. 1071.

MOTEGI, H., et al., TRAF6 negatively regulates the Jak1-Erk pathway in interleukin-2 signaling. Genes Cells, 2011. 16(2): p. 179-89.

MOOTHA, V.K., et al., PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet, 2003. 34(3): p. 267-73.

MORAN, C. J., ARENBERG, D. A., HUANG, C. C., GIORDANO, T. J., THOMAS, D. G., MISEK, D. E., CHEN, G., IANNETTONI, M. D., ORRINGER, M. B.,

HANASH, S. & BEER, D. G. 2002. RANTES expression is a predictor of survival in stage I lung adenocarcinoma. *Clin Cancer Res,* 8, 3803-12.

MU, C.Y., et al., High expression of PD-L1 in lung cancer may contribute to poor prognosis and tumor cells immune escape through suppressing tumor infiltrating dendritic cells maturation. Med Oncol, 2011. 28(3): p. 682-8.

NIKOLOVA, T., M. CHRISTMANN, and B. KAINA, FEN1 is overexpressed in testis, lung and brain tumors. Anticancer Res, 2009. 29(7): p. 2453-9.

OHRI, C. M., SHIKOTRA, A., GREEN, R. H., WALLER, D. A. & BRADDING, P. 2010. Chemokine receptor expression in tumour islets and stroma in non-small cell lung cancer. *BMC Cancer,* 10, 172.

OHTSUBO, M., et al., Human cyclin E, a nuclear protein essential for the G1-to-S phase transition. Mol Cell Biol, 1995. 15(5): p. 2612-24.

PAZOLLI, E., et al., Senescent stromal-derived osteopontin promotes preneoplastic cell growth. Cancer Res, 2009. 69(3): p. 1230-9.

PENG, C.Y., et al., Mitotic and G2 checkpoint control: regulation of 14-3-3 protein binding by phosphorylation of Cdc25C on serine-216. Science, 1997. 277(5331): p. 1501-5.

QIAO, X., et al., APC/C-Cdh1: from cell cycle to cellular differentiation and genomic integrity. Cell Cycle, 2010. 9(19): p. 3904-12.

RAYET, B. and C. GELINAS, Aberrant rel/nfkb genes and activity in human cancer. Oncogene, 1999. 18(49): p. 6938-47.

RILEY, J.L., PD-1 signaling in primary T cells. Immunol Rev, 2009. 229(1): p. 114-25.

SAHARIA, A., et al., Flap endonuclease 1 contributes to telomere stability. Curr Biol, 2008. 18(7): p. 496-500.

SAMARAJIWA, S.A., et al., INTERFEROME: the database of interferon regulated genes. Nucleic Acids Res, 2009. 37(Database issue): p. D852-7.

SANTAMARIA, D., et al., Cdk1 is sufficient to drive the mammalian cell cycle. Nature, 2007. 448(7155): p. 811-5.

SARENEVA, T., I. Julkunen, and S. Matikainen, IFN-alpha and IL-12 induce IL-18 receptor gene expression in human NK and T cells. J Immunol, 2000. 165(4): p. 1933-8.

SCHILLER, J.H., et al., Eastern cooperative group trial of interferon gamma in metastatic melanoma: an innovative study design. Clin Cancer Res, 1996. 2(1): p. 29-36.

SCHLUNS, K.S., T. STOKLASEK, and L. LEFRANCOIS, The roles of interleukin-15 receptor alpha: trans-presentation, receptor component, or both? Int J Biochem Cell Biol, 2005. 37(8): p. 1567-71.

SCHRODER, K., et al., Interferon-gamma: an overview of signals, mechanisms and functions. J Leukoc Biol, 2004. 75(2): p. 163-89.

SHANAHAN, F., et al., Cyclin E associates with BAF155 and BRG1, components of the mammalian SWI-SNF complex, and alters the ability of BRG1 to induce growth arrest. Mol Cell Biol, 1999. 19(2): p. 1460-9.

SHEVDE, L.A., et al., Osteopontin: an effector and an effect of tumor metastasis. Curr Mol Med, 2010. 10(1): p. 71-81.

SINGAL, D. P., YE, M. & QIU, X. 1996. Molecular basis for lack of expression of HLA class I antigens in human small-cell lung carcinoma cell lines. *Int J Cancer,* 68, 629-36.

SPRUCK, C.H., K.A. WON, and S.I. REED, Deregulated cyclin E induces chromosome instability. Nature, 1999. 401(6750): p. 297-300.

STONE, G. W., BARZEE, S., SNARSKY, V., SANTUCCI, C., TRAN, B., LANGER, R., ZUGATES, G. T., ANDERSON, D. G. & KORNBLUTH, R. S. 2009. Nanoparticle-delivered multimeric soluble CD40L DNA combined with Toll-Like Receptor agonists as a treatment for melanoma. *PLoS One,* 4, e7334.

SUBRAMANIAN, A., TAMAYO, P., MOOTHA, V. K., MUKHERJEE, S., EBERT, B. L., GILLETTE, M. A., PAULOVICH, A., POMEROY, S. L., GOLUB, T. R., LANDER, E. S., MESIROV, J. P., 2005, "Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles", *Proc Natl Acad Sci, USA*,102:43, 15545-15550.

SWANN, J. B., HAYAKAWA, Y., ZERAFA, N., SHEEHAN, K. C., SCOTT, B., SCHREIBER, R. D., HERTZOG, P. & SMYTH, M. J. 2007. Type I IFN contributes to NK cell homeostasis, activation, and antitumor function. *J Immunol,* 178, 7540-9.

TAKEUCHI, E., et al., Induction by interleukin-15 of human killer cell activity against lung cancer cell lines and its regulatory mechanisms. Jpn J Cancer Res, 1996. 87(12): p. 1251-8.

TAMM, I., et al., IAP-family protein survivin inhibits caspase activity and apoptosis induced by Fas (CD95), Bax, caspases, and anticancer drugs. Cancer Res, 1998. 58(23): p. 5315-20.

TANG, L., et al., Transcriptional inhibition of p21WAF1/CIP1 gene (CDKN1) expression by survivin is at least partially p53-dependent: evidence for survivin acting as a transcription factor or co-factor. Biochem Biophys Res Commun, 2012. 421(2): p. 249-54.

TANAKA, S., HATTORI, N., ISHIKAWA, N., HORIMASU, Y., DEGUCHI, N., TAKANO, A., TOMODA, Y., YOSHIOKA, K., FUJITAKA, K., ARIHIRO, K., OKADA, M., YOKOYAMA, A. & KOHNO, N. 2012. Interferon (alpha, beta and omega) receptor 2 is a prognostic biomarker for lung cancer. *Pathobiology,* 79, 24-33.

TANG, F., ZHAO, L. T., JIANG, Y., BA DE, N., CUI, L. X. & HE, W. 2008. Activity of recombinant human interleukin-15 against tumor recurrence and metastasis in mice. *Cell Mol Immunol,* 5, 189-96.

TEAGUE, R.M., et al., Interleukin-15 rescues tolerant CD8+ T cells for use in adoptive immunotherapy of established tumors. Nat Med, 2006. 12(3): p. 335-41.

TSAI, M. F., WANG, C. C., CHANG, G. C., CHEN, C. Y., CHEN, H. Y., CHENG, C. L., YANG, Y. P., WU, C. Y., SHIH, F. Y., LIU, C. C., LIN, H. P., JOU, Y. S., LIN, S. C., LIN, C. W., CHEN, W. J., CHAN, W. K., CHEN, J. J. & YANG, P. C. 2006. A new tumor suppressor DnaJ-like heat shock protein, HLJ1, and survival of patients with non-small-cell lung carcinoma. *J Natl Cancer Inst,* 98, 825-38.

VESELY, M.D., et al., Natural innate and adaptive immunity to cancer. Annu Rev Immunol, 2011. 29: p. 235-71.

WANG, S., RAVEN, J. F., DURBIN, J. E. & KOROMILAS, A. E. 2008. Stat1 phosphorylation determines Ras oncogenicity by regulating p27 kip1. *PLoS One,* 3, e3476.

WIKMAN, H. and E. KETTUNEN, Regulation of the G1/S phase of the cell cycle and alterations in the RB pathway in human lung cancer. Expert Rev Anticancer Ther, 2006. 6(4): p. 515-30.

WINTER, H., VAN DEN ENGEL, N. K., RUTTINGER, D., SCHMIDT, J., SCHILLER, M., POEHLEIN, C. H., LOHE, F., FOX, B. A., JAUCH, K. W., HATZ, R. A. & HU, H. M. 2007. Therapeutic T cells induce tumor-directed chemotaxis of innate immune cells through tumor-specific secretion of chemokines and stimulation of B16BL6 melanoma to secrete chemokines. *J Transl Med,* 5, 56.

YOU, Z., et al., Roles of Mcm7 and Mcm4 subunits in the DNA helicase activity of the mouse Mcm4/6/7 complex. J Biol Chem, 2002. 277(45): p. 42471-9.

YU, H. and R. Jove, The STATs of cancer--new molecular targets come of age. Nat Rev Cancer, 2004. 4(2): p. 97-105.

ZAIDI, M.R. AND G. MERLINO, The two faces of interferon-gamma in cancer. Clin Cancer Res, 2011. 17(19): p. 6118-24.

ZHENG, D. L., HUANG, Q. L., ZHOU, F., HUANG, Q. J., LIN, J. Y. & LIN, X. 2012. PA28beta regulates cell invasion of gastric cancer via modulating the expression of chloride intracellular channel 1. *J Cell Biochem,* 113, 1537-46.

CLAIMS

1. A method of providing a prognosis on the development of lung cancer in an individual, said method comprising the steps of

   i. determining in a sample taken from an individual, whose prognosis on lung cancer is to be provided, a test value for the expression of a biomarker; and

   ii. determining in a sample taken from a non-smoker suffering from lung cancer and/or from a smoker suffering from lung cancer a reference value for the expression of a biomarker, and

   iii. comparing the test values obtained for each biomarker with the reference values obtained from a non-smoker suffering from lung cancer and/or a smoker suffering from lung cancer,

   wherein said biomarker corresponds to at least 2 of the genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally, in combination with at least 1 and up to 108 genes, which are different from the above identified genes and are depicted up to 108 genes, which are different from the above identified genes and are depicted depicted in table 1, table 2, or both table 1 and table 2.

2. The method of claim 1 of providing a prognosis on the development of lung cancer in an individual, said method comprising the steps of

   i. determining in a sample taken from an individual, whose prognosis on lung cancer is to be provided, a test value of the expression of a biomarker comprising using a composition, comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide corresponding to at least 2 of the genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1), and, optionally, further comprising in combination with at least 1 and up to 108 different isolated nucleic acid

molecules each of which comprising a biomarker polynucleotide corresponding to at least 1 and up to 108 genes, which are different from the above identified genes and are depicted in table 1, table 2, or both table 1 and table 2; and

ii. determining a reference value in a sample taken from a non-smoker suffering from lung cancer and/or a smoker suffering from lung cancer comprising using a composition comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide corresponding to the genes identified in section i); and

iii. comparing the test values obtained for each biomarker in the composition, with reference values for the biomarkers from lung cancer of a non-smoker and/or a smoker suffering from lung cancer.

3. A method of providing a prognosis on the development of lung cancer in an individual, said method comprising the steps of

i. determining in a sample taken from an individual, whose prognosis on lung cancer is to be provided, a test value of the expression of a biomarker comprising using a composition, comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide corresponding to the genes identified by GeneBank Accession numbers 204767_s_at (FEN1) and 209545_s_at (RIPK2) or 204767_s_at (FEN1) and 213523_at (CCNE1) or 209545_s_at (RIPK2) and 213523_at (CCNE1), and, optionally, further comprising at least 1 and up to 108 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide corresponding to at least 1 and up to 108 genes, which are different from the above identified genes and are depicted depicted in table 1, table 2, or both table 1 and table 2; and

ii. determining a reference value in a sample taken from a non-smoker suffering from lung cancer and/or a smoker suffering from lung cancer comprising using a composition comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide corresponding to the genes identified in section i); and

iii. comparing the values obtained for each biomarker in the composition, with reference values for the biomarkers from lung cancer of a non-smoker and/or a smoker.

4. A method of providing a prognosis on the development of lung cancer in an individual, said method comprising the steps of

i. determining in a sample taken from an individual, whose prognosis on lung cancer is to be provided, a test value of the expression of a biomarker comprising using a composition, comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide corresponding to the genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1), and, optionally, further comprising at least 1 and up to 108 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide corresponding to at least 1 and up to 108 genes, which are different from the above identified genes and are depicted depicted in table 1, table 2, or both table 1 and table 2; and

ii. determining a reference value in a sample taken from a non-smoker suffering from lung cancer and/or a smoker suffering from lung cancer comprising using a composition comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide corresponding to the genes identified in section i); and

iii. comparing the values obtained for each biomarker in the composition, with reference values for the biomarkers from lung cancer of a non-smoker and/or a smoker.

5. The method of any one of claims 1 to 4, wherein differential expression is determined by way of comparison to a reference value obtained from an individual, which is a non-smoker and/or a smoker, who suffers from lung cancer or from a population of non-smokers and/or smokers who have lung cancer.

6. The method of any of the preceding claims, wherein a difference or a similarity in the test values and the reference values is used to forecast or classify the subject into a poor survival group or a good survival group.

7. The method of claim 6, wherein a test value which is (a) similar to the reference value obtained for a smoker with lung cancer signifies a poor prognosis, whereas a test value which is (b) similar to the reference value obtained for a non-smoker with lung cancer signifies an improved prognosis.

8. A method for predicting or monitoring the outcome of a lung cancer treatment in an individual suffering from lung cancer comprising

   i. determining in a biological sample of an individual who suffers from lung cancer a test value for the expression of each of at least 2 of the genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1) optionally in combination with at least 1 and up to 108 different biomarker genes depicted in table 1, table 2, or both table 1 and table 2;

   ii. determining in a biological sample of a control individual a reference value for the expression of each of the same biomarker genes used in step (i); and

   iii. comparing the values obtained for each said biomarker genes with the reference value;

   wherein said control individual is a non-smoker suffering from lung cancer, wherein differences between the test values and the reference values indicate the probability of said individual's lung cancer sharing one or more treatment outcomes of smoking-related lung cancer.

9. A method for predicting or monitoring the outcome of a lung cancer treatment in an individual suffering from lung cancer comprising

   i. determining in a biological sample of an individual who suffers from lung cancer a test value for the expression of the genes identified by GeneBank Accession numbers 204767_s_at (FEN1) and 209545_s_at (RIPK2), or 204767_s_at (FEN1) and 213523_at (CCNE1), or 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally in combination with at least 1 and up to 108 different biomarker genes depicted in table 1, table 2, or both table 1 and table 2;

ii. determining in a biological sample of a control individual a reference value for the expression of each of the same biomarker genes used in step (i); and

iii. comparing the values obtained for each said biomarker genes with the reference value;

wherein said control individual is a non-smoker suffering from lung cancer, wherein differences between the test values and the reference values indicate the probability of said individual's lung cancer sharing one or more treatment outcomes of smoking-related lung cancer.

10. A method for predicting or monitoring the outcome of a lung cancer treatment in an individual suffering from lung cancer comprising

i. determining in a biological sample of an individual who suffers from lung cancer a test value for the expression of each of the genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1) optionally in combination with at least 1 and up to 108 different biomarker genes depicted in table 1, table 2, or both table 1 and table 2;

ii. determining in a biological sample of a control individual a reference value for the expression of each of the same biomarker genes used in step (i); and

iii. comparing the values obtained for each said biomarker genes with the reference value;

wherein said control individual is a non-smoker suffering from lung cancer, wherein differences between the test values and the reference values indicate the probability of said individual's lung cancer sharing one or more treatment outcomes of smoking-related lung cancer.

11. A method for monitoring the progress of a lung cancer treatment in an individual, said method comprising determining at suitable time intervals before, during, or after lung cancer therapy, particularly at different time points during the treatment, in a sample taken from said individual, differential expression of biomarker genes comprising using a composition, comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide that

corresponds to at least 2 of the genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1) optionally in combination with at least 1 and up to 108 different biomarkers selected from the biomarkers depicted in Table 1, Table 2, or both Table 1 and Table 2.

12. A method for monitoring the progress of a lung cancer treatment in an individual, said method comprising determining at suitable time intervals before, during, or after lung cancer therapy, particularly at different time points during the treatment, in a sample taken from said individual, differential expression of biomarker genes comprising using a composition, comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide that corresponds to a gene identified by GeneBank Accession numbers 204767_s_at (FEN1) and 209545_s_at (RIPK2), or 204767_s_at (FEN1) and 213523_at (CCNE1), or 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally in combination with at least 1 and up to 108 different biomarker genes depicted in table 1, table 2, or both table 1 and table 2.

13. A method for monitoring the progress of a lung cancer treatment in an individual, said method comprising determining at suitable time intervals before, during, or after lung cancer therapy, particularly at different time points during the treatment, in a sample taken from said individual, differential expression of biomarker genes comprising using a composition, comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide that corresponds to a gene identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1) optionally in combination with at least 1 and up to 108 different biomarkers selected from the biomarkers depicted in Table 1, Table 2, or both Table 1 and Table 2.

14. The method of claim 13, comprising

    i. determining in a plurality of biological samples a value for the expression of each of the at least 2 of the genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1) optionally in combination with 1 and up to 108 different biomarker genes depicted in table 1, table 2, or both table 1 and table 2, wherein said plurality of biological samples are obtained at a plurality of

time points from an individual having lung cancer and receiving a lung cancer therapy;

ii. comparing the values obtained for each said biomarker genes with values of each said biomarker gene obtained at another time point and optionally with reference values for each said biomarker gene of a non-smoker suffering from lung cancer;

wherein differences between the test values obtained at the plurality of time points, or changes in the differences between the test values and the reference values provide a prediction of the outcome of the lung cancer therapy.

15. The method of claim 14, comprising

i. determining in a plurality of biological samples a value for the expression of each of the genes identified by GeneBank Accession numbers 204767_s_at (FEN1) and 209545_s_at (RIPK2), or 204767_s_at (FEN1) and 213523_at (CCNE1), or 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally in combination with at least 1 and up to 108 different biomarker genes depicted in table 1, table 2, wherein said plurality of biological samples are obtained at a plurality of time points from an individual having lung cancer and receiving a lung cancer therapy;

ii. comparing the values obtained for each said biomarker genes with values of each said biomarker gene obtained at another time point and optionally with reference values for each said biomarker gene of a non-smoker suffering from lung cancer;

wherein differences between the test values obtained at the plurality of time points, or changes in the differences between the test values and the reference values provide a prediction of the outcome of the lung cancer therapy.

16. The method of claim 15, comprising

i. determining in a plurality of biological samples a value for the expression of each of the genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1) optionally in combination with 1 and up to 108 different biomarker genes depicted in table 1, table 2, or both table 1 and table 2, wherein said plurality of biological samples are obtained at a plurality of time points

from an individual having lung cancer and receiving a lung cancer therapy;

ii. comparing the values obtained for each said biomarker genes with values of each said biomarker gene obtained at another time point and optionally with reference values for each said biomarker gene of a non-smoker suffering from lung cancer;

wherein differences between the test values obtained at the plurality of time points, or changes in the differences between the test values and the reference values provide a prediction of the outcome of the lung cancer therapy.

17. The method according to any one of the preceding claims, wherein the sample is selected from blood, serum, plasma, sputum, saliva, tissue particularly lung tissue, obtained through biopsy, bronchia brushings, exhaled breath, or urine.

18. The method according to any one of the preceding claims, wherein determination of biomarker values is accomplished by performing an *in-vitro* assay, particularly an *in-vitro* assay selected from the group consisting of an antibody-based assay such as an immunoassay, a histological or cytological assay, an expression level assay such as an RNA expression level assay and an aptamer-based assay.

19. The method according to any one of the preceding claims, wherein

a. a polynucleotide or a variant thereof, which is complementary to a target gene as depicted in table 1, table 2, or both table 1 and table 2, is used as a molecular probe in a hybridization reaction or as a molecular primer in a nucleic acid extension reaction, for the determination of the target and reference value.

b. one or more detectably labeled antibodies are used in the determination of the target and reference value, which antibodies are capable of identifying biomarker gene products encoded by one or more biomarker genes depicted in table 1, table 2, or both table 1 and table 2, or by conserved variants or peptide fragments thereof.

20. The method according to any one of the preceding claims, wherein at least 2 of the genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1) and optionally in combination with

at least 1 biomarker or a composition, is used comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide that corresponds to a different gene

(a) selected from the group of biomarker genes depicted in table 1 identified by GeneBank Accession number 200887_s_at (STAT1), 204785_x_at (IFNAR2); 201762_s_at (PSME2), 204279_at (PSMB9), 203915_at (CXCL9), 204533_at (CXCL10), 227458_at (CD274), 207375_s_at (IL15), 207375_s_at (IL15RA), 205207_at (IL6), 204655_at (CCL5), 204103_at (CCL4), 207072_at (IL18RAP), and 205291_at (IL2RB); or

(b) selected from the group of biomarker genes depicted in table 1 identified by GeneBank Accession number 203213_at (CDK1), 222036_s_at (MCM4), 202870_s_at (CDC20); 202095_s_at (BIRC5), 202246_s_at (CDK4), and 207574_s_at (GADD45B); or

(c) selected from the group of biomarker genes depicted in table 1 and table 2 identified by GeneBank Accession number 208373_s_at (P2RY6), 210563_x_at (CFLAR), 210538_s_at (BIRC3), 207535_s_at (NFKB2), 203868_s_at (CHEK1), 203666_at (CXCL12), 209875_s_at (SPP1), 207697_x_at (LILRB2), 204639_at (ADA), and 224818_at (SORT1).

21. The method according to any one of the preceding claims, wherein 204767_s_at (FEN1) and 209545_s_at (RIPK2), or 204767_s_at (FEN1) and 213523_at (CCNE1), or 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally in combination with at least 1 biomarker or a composition is used comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide that corresponds to a different gene

(a) selected from the group of biomarker genes depicted in table 1 identified by GeneBank Accession number 200887_s_at (STAT1), 204785_x_at (IFNAR2); 201762_s_at (PSME2), 204279_at (PSMB9), 203915_at (CXCL9), 204533_at (CXCL10), 227458_at (CD274), 207375_s_at (IL15), 207375_s_at (IL15RA), 205207_at (IL6), 204655_at (CCL5), 204103_at (CCL4), 207072_at (IL18RAP), and 205291_at (IL2RB); or

(b) selected from the group of biomarker genes depicted in table 1 identified by GeneBank Accession number 203213_at (CDK1), 222036_s_at (MCM4), 202870_s_at (CDC20); 202095_s_at (BIRC5), 202246_s_at (CDK4), and 207574_s_at (GADD45B); or

(c) selected from the group of biomarker genes depicted in table 1 and table 2 identified by GeneBank Accession number 208373_s_at (P2RY6), 210563_x_at (CFLAR), 210538_s_at (BIRC3), 207535_s_at (NFKB2), 203868_s_at (CHEK1), 203666_at (CXCL12), 209875_s_at (SPP1), 207697_x_at (LILRB2), 204639_at (ADA), and 224818_at (SORT1).

22. The method according to any one of the preceding claims, wherein the genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1) and optionally in combination with at least 1 biomarker or a composition, is used comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide that corresponds to a different gene

(a) selected from the group of biomarker genes depicted in table 1 identified by GeneBank Accession number 200887_s_at (STAT1), 204785_x_at (IFNAR2); 201762_s_at (PSME2), 204279_at (PSMB9), 203915_at (CXCL9), 204533_at (CXCL10), 227458_at (CD274), 207375_s_at (IL15), 207375_s_at (IL15RA), 205207_at (IL6), 204655_at (CCL5), 204103_at (CCL4), 207072_at (IL18RAP), and 205291_at (IL2RB); or

(b) selected from the group of biomarker genes depicted in table 1 identified by GeneBank Accession number 203213_at (CDK1), 222036_s_at (MCM4), 202870_s_at (CDC20); 202095_s_at (BIRC5), 202246_s_at (CDK4), and 207574_s_at (GADD45B); or

(c) selected from the group of biomarker genes depicted in table 1 and table 2 identified by GeneBank Accession number 208373_s_at (P2RY6), 210563_x_at (CFLAR), 210538_s_at (BIRC3), 207535_s_at (NFKB2), 203868_s_at (CHEK1), 203666_at (CXCL12), 209875_s_at (SPP1), 207697_x_at (LILRB2), 204639_at (ADA), and 224818_at (SORT1).

23. A composition comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide that corresponds to at least 2 of the genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1) and optionally in combination with at least 1 and up to 108 different biomarker genes depicted in table 1, table 2, or both table 1 and table 2.

24. A composition comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide that corresponds to biomarker genes 204767_s_at (FEN1) and 209545_s_at (RIPK2), or 204767_s_at (FEN1) and 213523_at (CCNE1), or 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally in combination with at least 1 and up to 108 different biomarker genes depicted in table 1, table 2, or both table 1 and table 2.

25. A composition comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide that corresponds to biomarker genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1) and optionally in combination with at least 1 and up to 108 different biomarker genes depicted in table 1, table 2, or both table 1 and table 2.

26. The composition of any one off the preceding claims, wherein said biomarker polynucleotide is
    a. a polynucleotide or a variant thereof, which is complementary to a target gene as depicted in table 1, table 2, or both table 1 and table 2 and can be used as a hybridization probe or a primer; or
    b. one or more detectably labeled antibodies, which antibodies are capable of identifying biomarker gene products encoded by one or more biomarker genes depicted in table 1, table 2, or both table 1 and table 2, or by conserved variants or peptide fragments thereof.

27. The composition of any one of the preceding claims, wherein said genes are at least 2 of the genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally in combination with at least 1 biomarker gene
    (a) selected from the group of biomarker genes depicted in table 1 identified by GeneBank Accession number 200887_s_at (STAT1), 204785_x_at (IFNAR2); 201762_s_at (PSME2), 204279_at (PSMB9), 203915_at (CXCL9), 204533_at (CXCL10), 227458_at (CD274), 207375_s_at (IL15), 207375_s_at (IL15RA), 205207_at (IL6), 204655_at (CCL5), 204103_at (CCL4), 207072_at (IL18RAP), and 205291_at (IL2RB); or
    (b) selected from the group of biomarker genes depicted in table 1 identified by GeneBank Accession number 203213_at (CDK1),, 222036_s_at

(MCM4), 202870_s_at (CDC20); 202095_s_at (BIRC5), 202246_s_at (CDK4), and 207574_s_at (GADD45B); or

(c) selected from the group of biomarker genes depicted in table 1 and table 2 identified by GeneBank Accession number 208373_s_at (P2RY6), 210563_x_at (CFLAR), 210538_s_at (BIRC3), 207535_s_at (NFKB2), 203868_s_at (CHEK1), 203666_at (CXCL12), 209875_s_at (SPP1), 207697_x_at (LILRB2), 204639_at (ADA), and 224818_at (SORT1).

28. The composition of any one of the preceding claims, wherein said genes are the genes identified by GeneBank Accession numbers 204767_s_at (FEN1) and 209545_s_at (RIPK2), or 204767_s_at (FEN1) and 213523_at (CCNE1), or 209545_s_at (RIPK2) and 213523_at (CCNE1), and optionally in combination with at least 1 biomarker gene

(a) selected from the group of biomarker genes depicted in table 1 identified by GeneBank Accession number 200887_s_at (STAT1), 204785_x_at (IFNAR2); 201762_s_at (PSME2), 204279_at (PSMB9), 203915_at (CXCL9), 204533_at (CXCL10), 227458_at (CD274), 207375_s_at (IL15), 207375_s_at (IL15RA), 205207_at (IL6), 204655_at (CCL5), 204103_at (CCL4), 207072_at (IL18RAP), and 205291_at (IL2RB); or

(b) selected from the group of biomarker genes depicted in table 1 identified by GeneBank Accession number 203213_at (CDK1),, 222036_s_at (MCM4), 202870_s_at (CDC20); 202095_s_at (BIRC5), 202246_s_at (CDK4), and 207574_s_at (GADD45B); or

(c) selected from the group of biomarker genes depicted in table 1 and table 2 identified by GeneBank Accession number 208373_s_at (P2RY6), 210563_x_at (CFLAR), 210538_s_at (BIRC3), 207535_s_at (NFKB2), 203868_s_at (CHEK1), 203666_at (CXCL12), 209875_s_at (SPP1), 207697_x_at (LILRB2), 204639_at (ADA), and 224818_at (SORT1).

29. The composition of any one of the preceding claims, wherein said genes are the genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally in combination with at least 1 gene

(a) selected from the group of biomarker genes depicted in table 1 identified by GeneBank Accession number 200887_s_at (STAT1), 204785_x_at

(IFNAR2); 201762_s_at (PSME2), 204279_at (PSMB9), 203915_at (CXCL9), 204533_at (CXCL10), 227458_at (CD274), 207375_s_at (IL15), 207375_s_at (IL15RA), 205207_at (IL6), 204655_at (CCL5), 204103_at (CCL4), 207072_at (IL18RAP), and 205291_at (IL2RB); or

(b) selected from the group of biomarker genes depicted in table 1 identified by GeneBank Accession number 203213_at (CDK1),, 222036_s_at (MCM4), 202870_s_at (CDC20); 202095_s_at (BIRC5), 202246_s_at (CDK4), and 207574_s_at (GADD45B); or

(c) selected from the group of biomarker genes depicted in table 1 and table 2 identified by GeneBank Accession number 208373_s_at (P2RY6), 210563_x_at (CFLAR), 210538_s_at (BIRC3), 207535_s_at (NFKB2), 203868_s_at (CHEK1), 203666_at (CXCL12), 209875_s_at (SPP1), 207697_x_at (LILRB2), 204639_at (ADA), and 224818_at (SORT1).

30. A kit for determining in an individual lung cancer or for monitoring the progress of a lung cancer treatment in an individual, comprising reagent for detecting differential expression of at least 2 of the biomarker genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally in combination with at least 1 and up to 108 genes, which are different from the above identified genes and are depicted depicted in table 1, table 2, or both table 1 and table 2.

31. A kit for determining in an individual lung cancer or for monitoring the progress of a lung cancer treatment in an individual, comprising reagent for detecting differential expression of a a biomarker genes identified by GeneBank Accession numbers 204767_s_at (FEN1) and 209545_s_at (RIPK2), or 204767_s_at (FEN1) and 213523_at (CCNE1), or 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally in combination with at least 1 and up to 108 genes, which are different from the above identified genes and are depicted depicted in table 1, table 2, or both table 1 and table 2.

32. A kit for determining in an individual lung cancer or for monitoring the progress of a lung cancer treatment in an individual, comprising reagent for detecting differential expression of a biomarker genes identified by GeneBank Accession numbers 204767_s_at (FEN1), 209545_s_at (RIPK2) and 213523_at (CCNE1), optionally in combination with at least 1 and up to 108 genes, which are different

from the above identified genes and are depicted depicted in table 1, table 2, or both table 1 and table 2.

33. A method of providing a prognosis on the development of lung cancer in an individual, said method comprising the steps of

    i. determining in a sample taken from an individual, whose prognosis on lung cancer is to be provided, a test value for the expression of a biomarker; and

    ii. determining in a sample taken from a non-smoker suffering from lung cancer and/or from a smoker suffering from lung cancer a reference value for the expression of a biomarker, and

    iii. comparing the test values obtained for each biomarker with the reference values obtained from a non-smoker suffering from lung cancer and/or a smoker suffering from lung cancer,

    wherein said biomarker corresponds to at least 2 and up to 111 genes depicted in table 1, table 2, or both table 1 and table 2.

34. A method for predicting or monitoring the outcome of a lung cancer treatment in an individual suffering from lung cancer comprising

    i. determining in a biological sample of an individual who suffers from lung cancer a test value for the expression of each of at least 2 and up to 111 different biomarker genes depicted in table 1, table 2, or both table 1 and table 2;

    ii. determining in a biological sample of a control individual a reference value for the expression of each of the same biomarker genes used in step (i); and

    iii. comparing the values obtained for each said biomarker genes with the reference value;

    wherein said control individual is a non-smoker suffering from lung cancer, wherein differences between the test values and the reference values indicate the probability of said individual's lung cancer sharing one or more treatment outcomes of smoking-related lung cancer.

35. A method for monitoring the progress of a lung cancer treatment in an individual, said method comprising determining at suitable time intervals before, during, or after lung cancer therapy, particularly at different time points during the treatment, in a sample taken from said individual, differential expression of biomarker genes comprising using a composition, comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide that corresponds to at least 2 and up to 111 different biomarkers selected from the biomarkers depicted in Table 1, Table 2, or both Table 1 and Table 2.

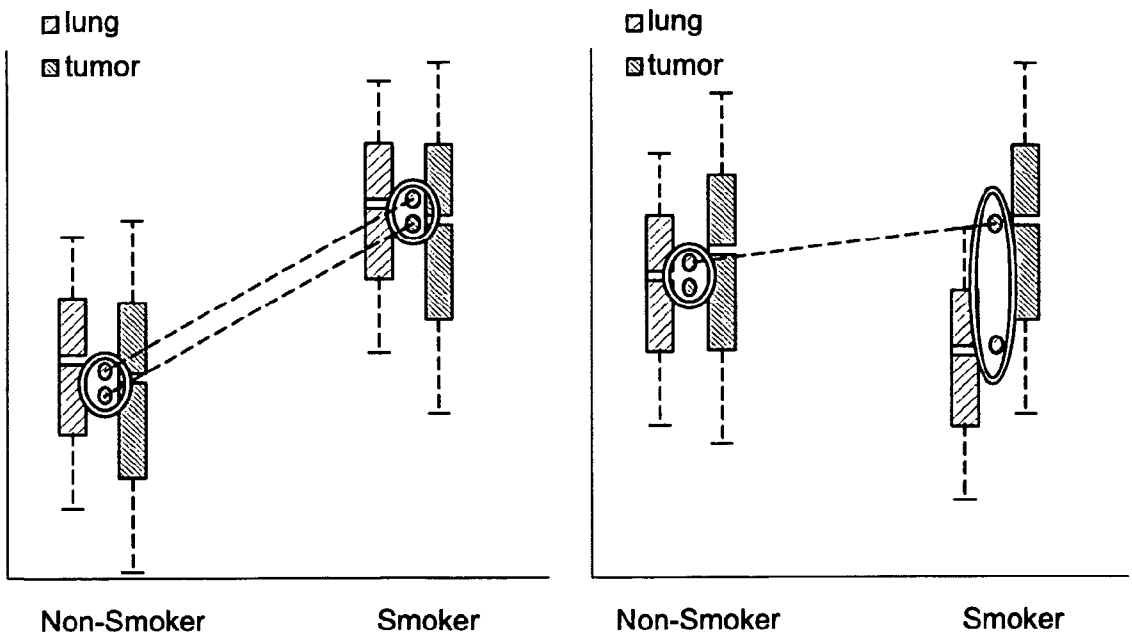


Fig. 1

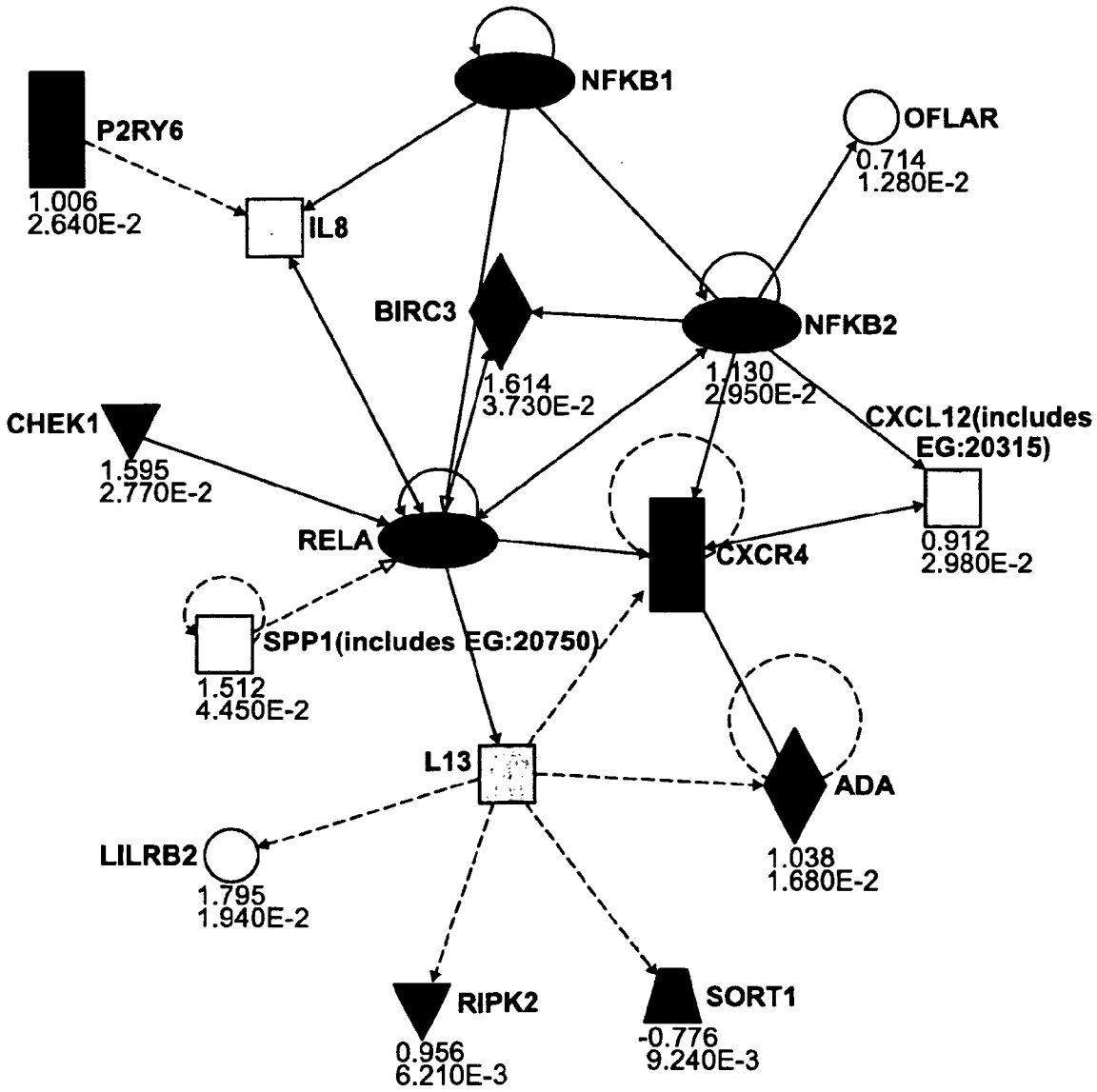

NFKB1
P2RY6
1.006
2.640E-2
IL8
OFLAR
0.714
1.280E-2
BIRC3
1.614
3.730E-2
NFKB2
1.130
2.950E-2
CHEK1
1.595
2.770E-2
CXCL12(includes EG:20315)
0.912
2.980E-2
RELA
CXCR4
SPP1(includes EG:20750)
1.512
4.450E-2
L13
ADA
1.038
1.680E-2
LILRB2
1.795
1.940E-2
RIPK2
0.956
6.210E-3
SORT1
-0.776
9.240E-3

Fig. 2

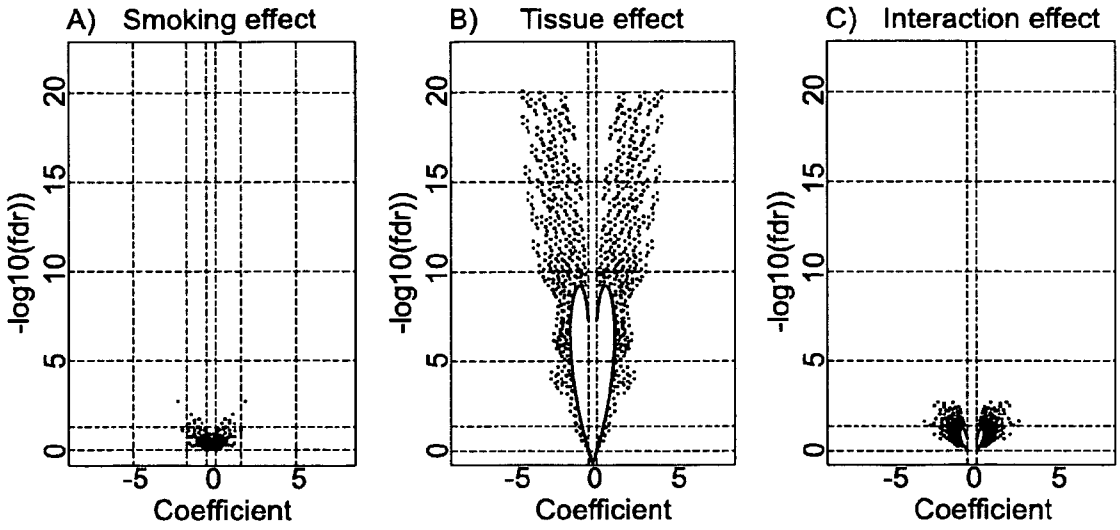
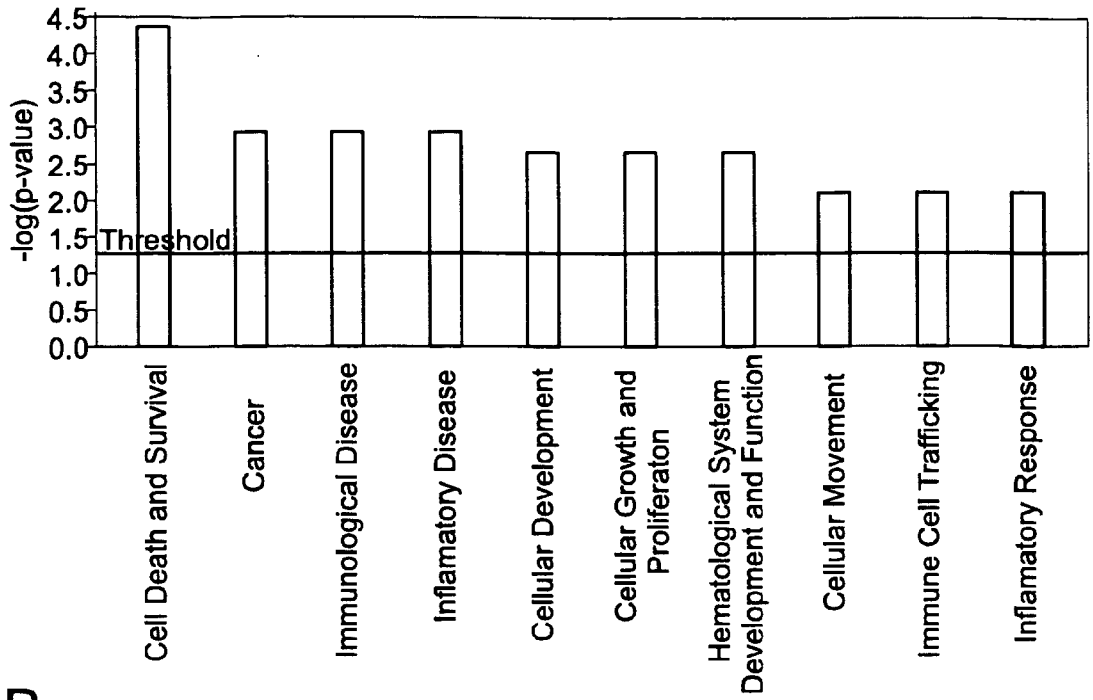

A) Smoking effect
B) Tissue effect
C) Interaction effect
-log10(fdr))
0
5
10
15
20
-5
0
5
Coefficient

Fig. 3


-log(p-value)
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Threshold
Cell Death and Survival
Cancer
Immunological Disease
Inflamatory Disease
Cellular Development
Cellular Growth and Proliferaton
Hematological System Development and Function
Cellular Movement
Immune Cell Trafficking
Inflamatory Response

B

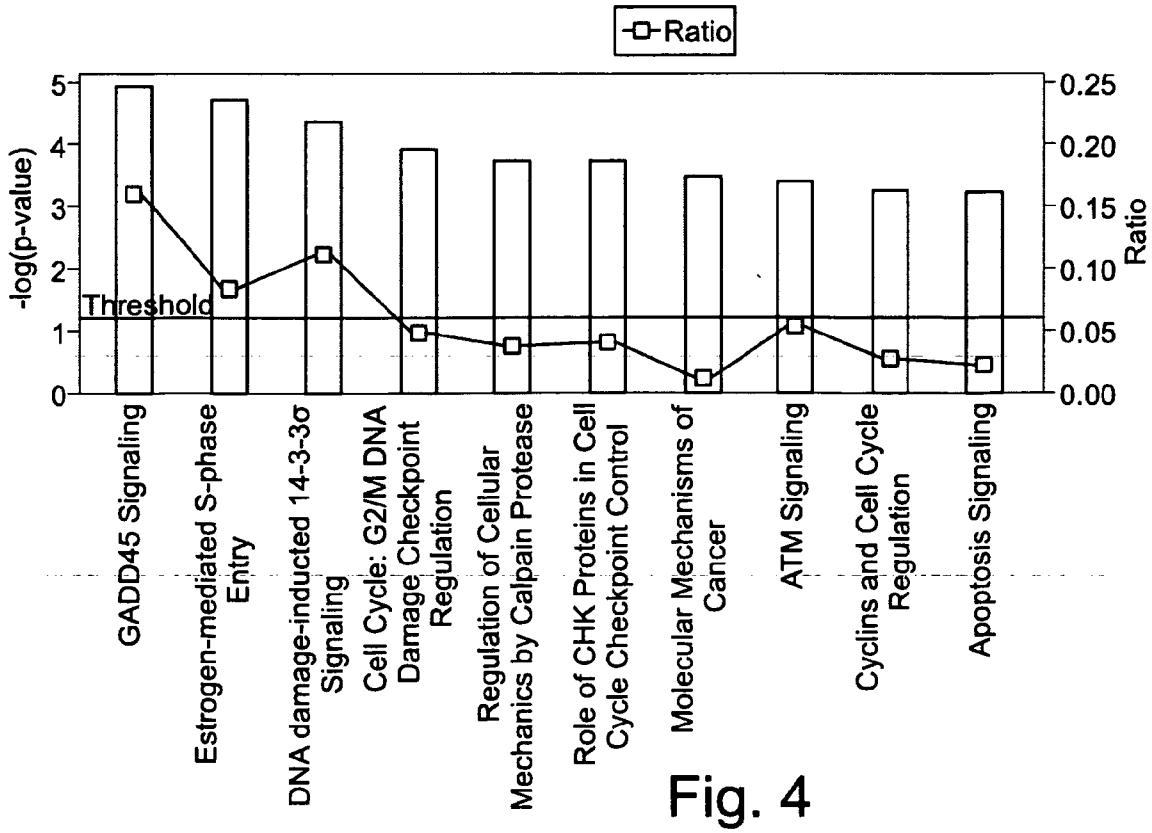

Ratio
-log(p-value)
Ratio
Threshold
GADD45 Signaling
Estrogen-mediated S-phase Entry
DNA damage-inducted 14-3-3σ Signaling
Cell Cycle: G2/M DNA Damage Checkpoint Regulation
Regulation of Cellular Mechanics by Calpain Protease
Role of CHK Proteins in Cell Cycle Checkpoint Control
Molecular Mechanisms of Cancer
ATM Signaling
Cyclins and Cell Cycle Regulation
Apoptosis Signaling

Fig. 4

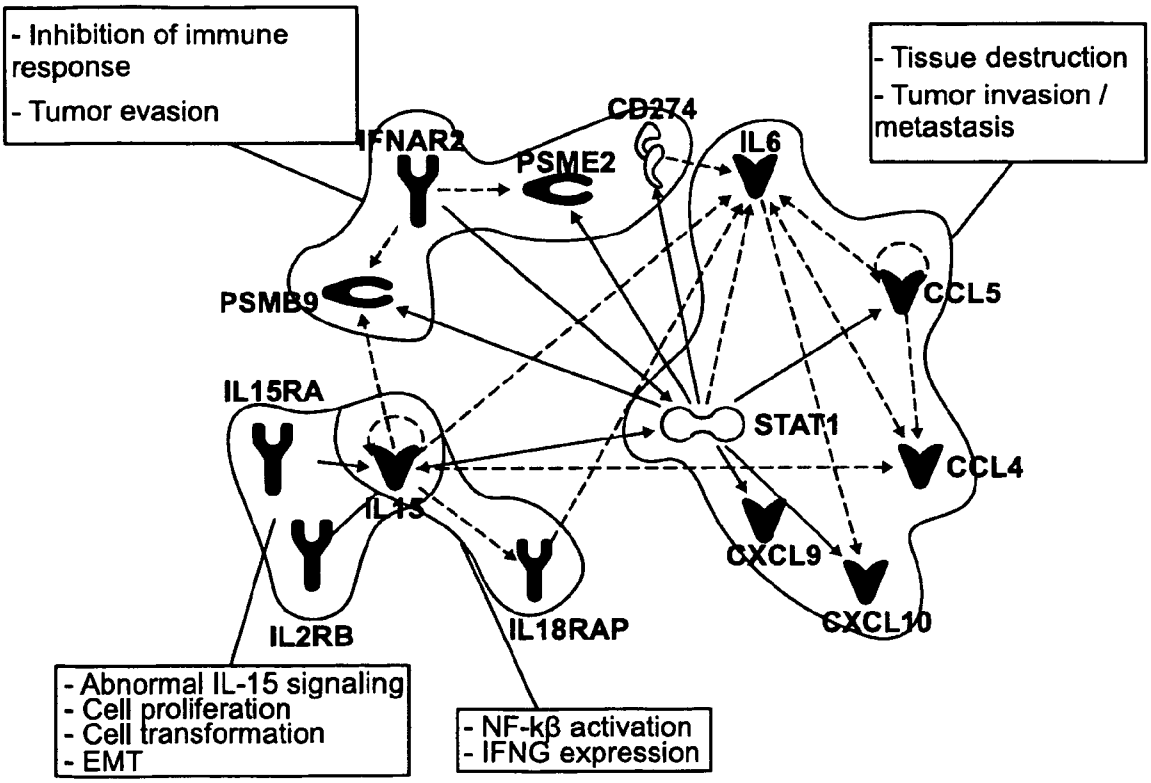

- Inhibition of immune response
- Tumor evasion
- Tissue destruction
- Tumor invasion / metastasis
IFNAR2
PSME2
CD274
IL6
PSMB9
CCL5
IL15RA
STAT1
CCL4
IL15
CXCL9
CXCL10
IL2RB
IL18RAP
- Abnormal IL-15 signaling
- Cell proliferation
- Cell transformation
- EMT
- NF-kβ activation
- IFNG expression

Fig. 5

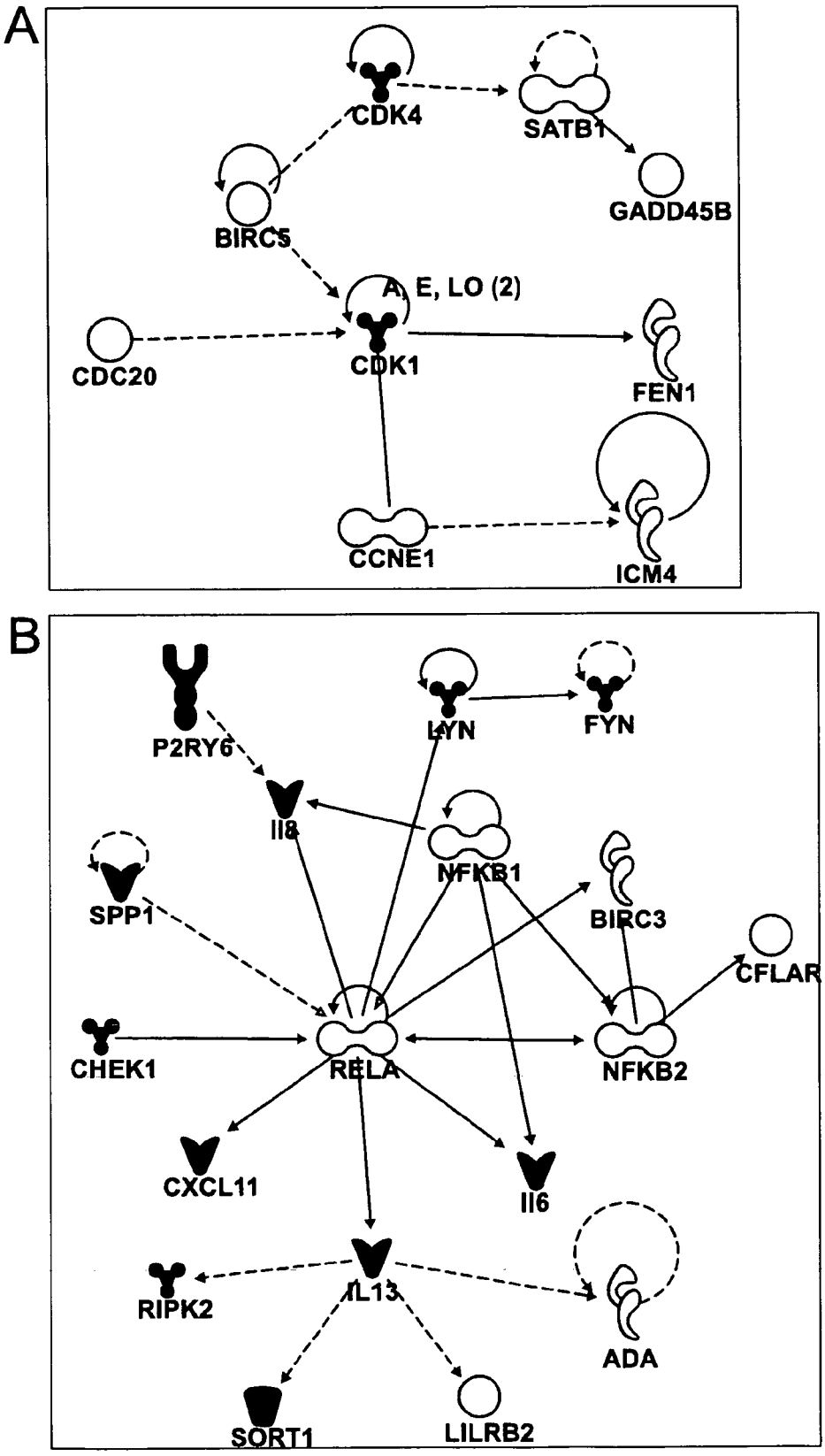

A
CDK4
SATB1
GADD45B
BIRC5
A, E, LO (2)
CDC20
CDK1
FEN1
CCNE1
ICM4
B
P2RY6
LYN
FYN
Il8
NFKB1
SPP1
BIRC3
CFLAR
CHEK1
RELA
NFKB2
CXCL11
Il6
RIPK2
IL13
ADA
SORT1
LILRB2

Fig. 6

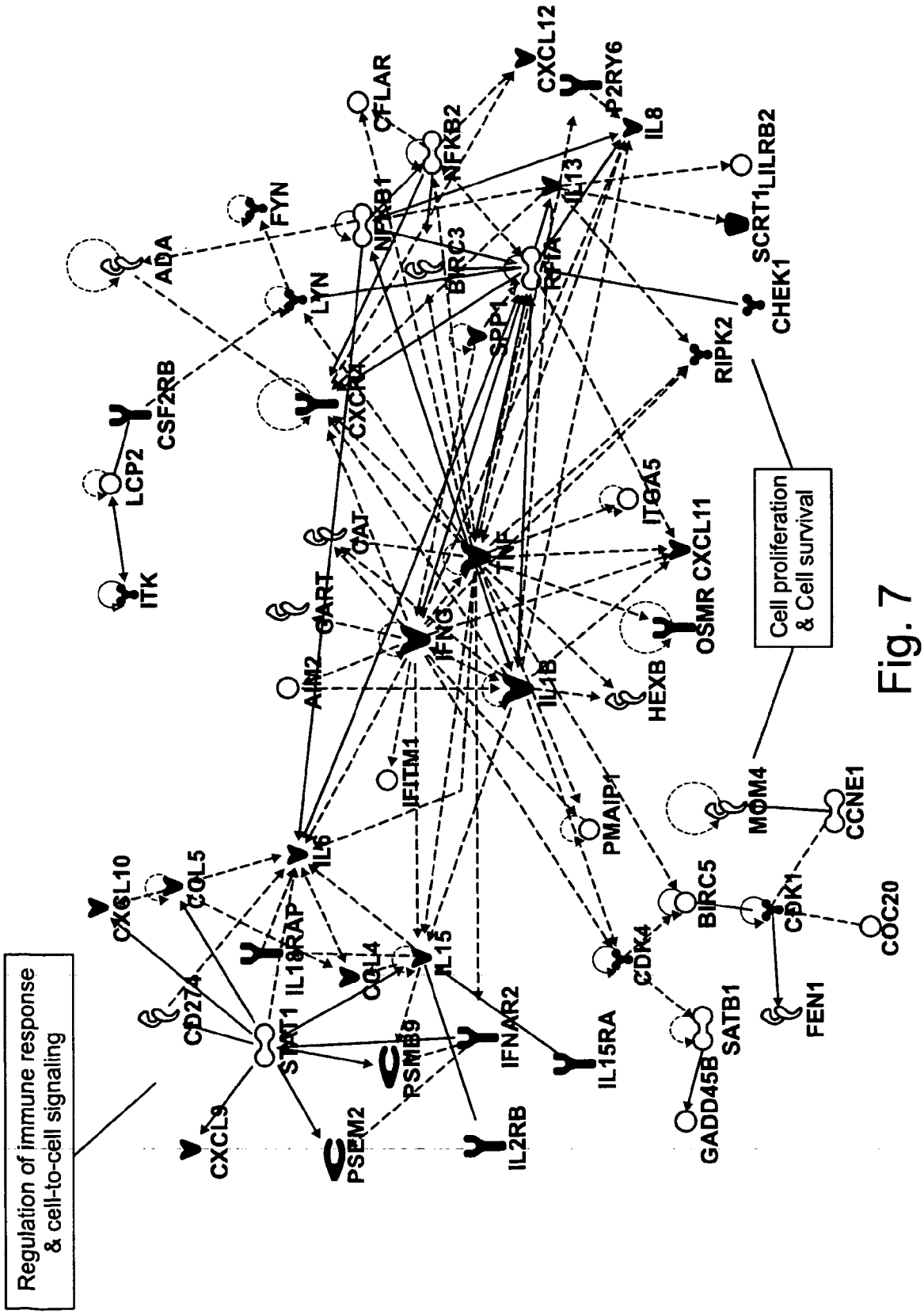

Regulation of immune response & cell-to-cell signaling
Cell proliferation & Cell survival
ADA
FYN
LYN
CSF2RB
LCP2
ITK
CFLAR
NFKB2
NFKB1
BIRC3
CXCL12
P2RY6
IL8
IL13
RFTA
SCRT1
LILRB2
CHEK1
RIPK2
SRP1
CXCR4
ITGA5
CXCL11
OSMR
HEXB
CAT
GART
AIM2
IFNG
TNF
IL1B
IFITM1
PMAIP1
IL6
CXCL10
CCL5
IL18RAP
CCL4
IL15
CD274
STAT1
PSMB9
IFNAR2
IL15RA
IL2RB
PSEM2
CXCL9
MCM4
CCNE1
BIRC5
CDK1
COC20
CDK4
SATB1
FEN1
GADD45B

Fig. 7

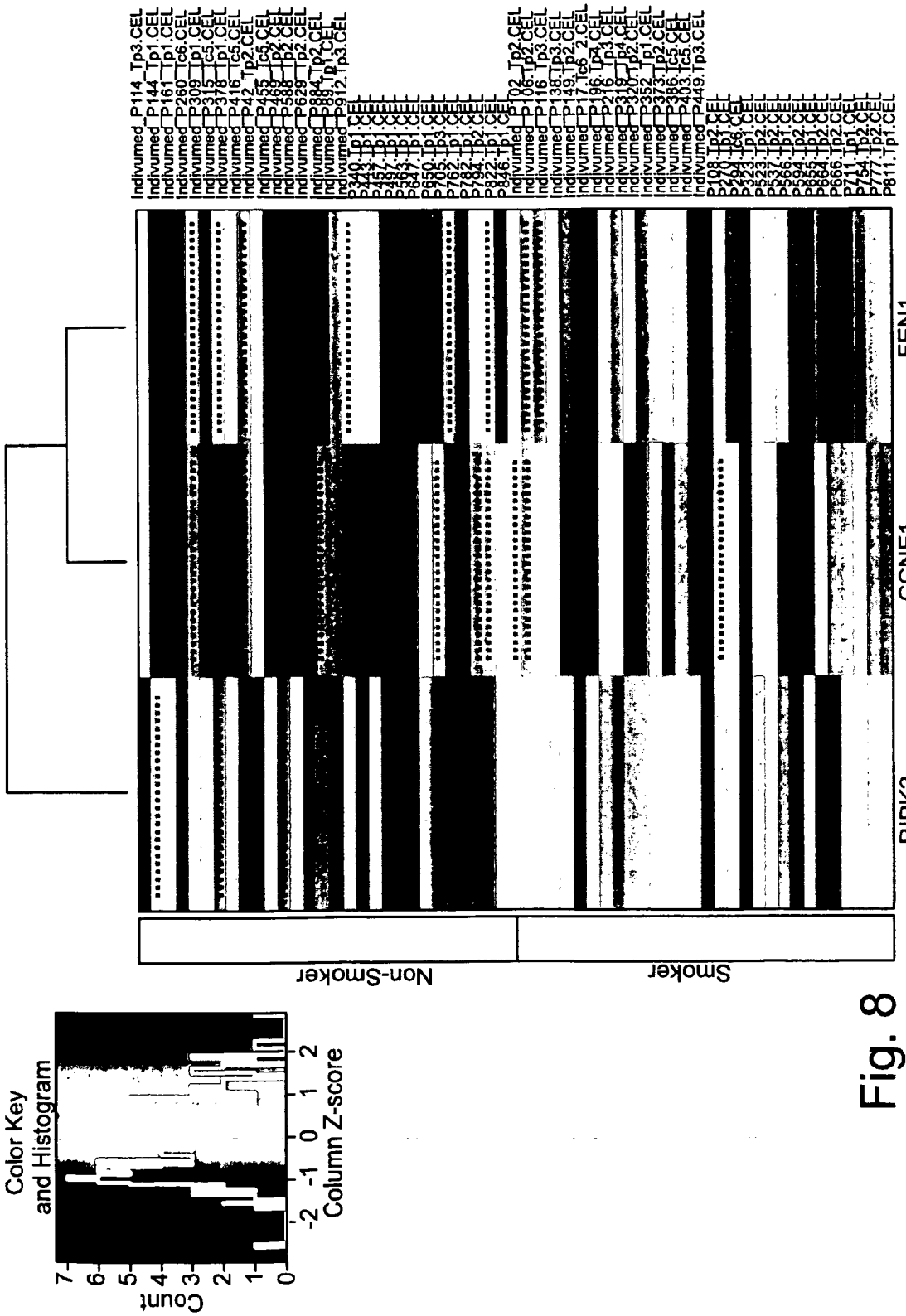

Indivumed_P114_Tp3.CEL
Indivumed_P144_Tp1.CEL
Indivumed_P161_Tp1.CEL
Indivumed_P260_Tc6.CEL
Indivumed_P309_Tp1.CEL
Indivumed_P312_Tc5.CEL
Indivumed_P378_Tp1.CEL
Indivumed_P416_Tc5.CEL
Indivumed_P42_Tp2.CEL
Indivumed_P455_Tc5.CEL
Indivumed_P469_Tp2.CEL
Indivumed_P588_Tp2.CEL
Indivumed_P629_Tp2.CEL
Indivumed_P884_Tp2.CEL
Indivumed_P889_Tp1.CEL
Indivumed_P912_Tp3.CEL
P340_Tp1.CEL
P443_Tp1.CEL
P457_Tp1.CEL
P563_Tp3.CEL
P847_Tp1.CEL
P650_Tp1.CEL
P705_Tp3.CEL
P782_Tp1.CEL
P794_Tp2.CEL
P822_Tp1.CEL
P846_Tp1.CEL
Indivumed_P102_Tp2.CEL
Indivumed_P106_Tp2.CEL
Indivumed_P116_Tp3.CEL
Indivumed_P138_Tp3.CEL
Indivumed_P149_Tp2.CEL
Indivumed_P17_Tc6_2.CEL
Indivumed_P196_Tp4.CEL
Indivumed_P216_Tp3.CEL
Indivumed_P319_Tp4.CEL
Indivumed_P320_Tp2.CEL
Indivumed_P352_Tp1.CEL
Indivumed_P373_Tp2.CEL
Indivumed_P386_Tc5.CEL
Indivumed_P403_Tc5.CEL
Indivumed_P449_Tp3.CEL
P108_Tp2.CEL
P270_Tp1.CEL
P294_Tc6.CEL
P523_Tp1.CEL
P537_Tp2.CEL
P566_Tp1.CEL
P594_Tp1.CEL
P664_Tp3.CEL
P666_Tp2.CEL
P711_Tp1.CEL
P754_Tp2.CEL
P777_Tp2.CEL
P811_Tp1.CEL
RIPK2
CCNE1
FEN1
Non-Smoker
Smoker
Color Key and Histogram
Count
Column Z-score

Fig. 8

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2013/060809

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68 G01N33/574
ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12Q 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, WPI Data, BIOSIS, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	MARIA TERESA LANDI ET AL: "Gene Expression Signature of Cigarette Smoking and Its Role in Lung Adenocarcinoma Development and Survival", PLOS ONE, vol. 3, no. 2, 1 January 2008 (2008-01-01) , pages e1651-e1651, XP055005066, ISSN: 1932-6203, DOI: 10.1371/journal.pone.0001651 page 3; tables 2,3 page 5 - page 6 ----- -/--	1-22, 33-35

[X] Further documents are listed in the continuation of Box C.

[X] 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	Date of mailing of the international search report
15 July 2013	24/09/2013

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

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	T. DUTU ET AL: "Differential expression of biomarkers in lung adenocarcinoma: a comparative study between smokers and never-smokers", ANNALS OF ONCOLOGY, vol. 16, no. 12, 1 December 2005 (2005-12-01), pages 1906-1914, XP055026741, ISSN: 0923-7534, DOI: 10.1093/annonc/mdi408 the whole document	1-22, 33-35
Y	WO 2010/099161 A1 (UNIV OHIO STATE RES FOUND [US]; US GOV HEALTH & HUMAN SERV [US]; CROCE) 2 September 2010 (2010-09-02) paragraph [0142] - paragraph [0145]; claims 13-18; figures 5,12; example 1; table 5	1-22, 33-35
X	AFFYMETRIX: "GeneChip TM Human Genome U133 Arrays", DATASHEET AFFYMETRIX,, 1 January 2003 (2003-01-01), pages 1-8, XP007921348, the whole document	23-32
Y	HUMAM KADARA ET AL: "Identification of Gene Signatures and Molecular Markers for Human Lung Cancer Prognosis using an In vitro Lung Carcinogenesis System", CANCER PREVENTION RESEARCH, XX, CN , vol. 2, no. 8 1 August 2009 (2009-08-01), pages 702-711, XP002679416, DOI: 10.1158/1940-6207.CAPR-09-0084 Retrieved from the Internet: URL:http://cancerpreventionresearch.aacrjournals.org/content/2/8/702 [retrieved on 2009-07-28] the whole document	1-22, 33-35
Y	JUNYA FUJIMOTO ET AL: "Comparative Functional Genomics Analysis of NNK Tobacco-Carcinogen Induced Lung Adenocarcinoma Development in Gprc5a-Knockout Mice", PLOS ONE, vol. 5, no. 7, 29 July 2010 (2010-07-29), page e11847, XP055071118, DOI: 10.1371/journal.pone.0011847 abstract	1-22, 33-35

-/--

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2013/060809

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	SATO MITSUO ET AL: "Increased expression and no mutation of the Flap endonuclease (FEN1) gene in human lung cancer", ONCOGENE, NATURE PUBLISHING GROUP, GB, vol. 22, no. 46, 16 October 2003 (2003-10-16), pages 7243-7246, XP002543517, ISSN: 0950-9232, DOI: 10.1038/SJ.ONC.1206977 abstract page 7245, column 1	1-22, 33-35
Y	C MÜLLER-TIDOW ET AL: "Cyclin E is the only cyclin-dependent kinase 2-associated cyclin that predicts metastasis and survival in early stage non-small cell lung cancer", CANCER RESEARCH, vol. 61, 1 January 2001 (2001-01-01), pages 647-653, XP055071192, the whole document page 652, column 1, paragraph 4 - column 2, paragraph 1	1-22, 33-35
Y	MISHINA T ET AL: "Cyclin E expression, a potential prognostic marker for non-small cell lung cancers", CLINICAL CANCER RESEARCH, THE AMERICAN ASSOCIATION FOR CANCER RESEARCH, US, vol. 6, no. 1, 1 January 2000 (2000-01-01) , pages 11-16, XP002619173, ISSN: 1078-0432 page 15, column 1 the whole document	1-22, 33-35

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2013/060809

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.:

10, 16, 22, 29, 32(completely); 1-9, 11-15, 17-21, 23-28, 30, 31
33-35(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: 10, 16, 22, 29, 32(completely); 1-9, 11-15, 17-21, 23-28, 30, 31, 33-35(partially)

Methods of providing a prognosis on the development of lung cancer in an individual, and methods for predicting or monitoring the outcome/progress of a lung cancer treatment in an individual, said methods comprising determining in a sample from an individual a test value of the expression of a panel of biomarkers corresponding to at least 2 and up to 111 different genes depicted in table 1, table 2, or both table 1 and table 2, wherein at least one of the biomarker genes is FEN1 , or wherein the markers compriseFEN1 and one or two selected fromRIPK2 andCCNE1 ,and comparing the values obtained for each biomarker with reference values obtained from lung cancer of a non-smoker and/or a smoker.
Compositions comprising comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide that corresponds to at least 2 of the biomarker genes FEN1, RIPK2 and CCNE1, wherein one of the genes is FEN1 , and optionally in combination with at least 1 and up to 108 different biomarker genes depicted in table 1, table 2, or both table 1 and table 2.
A kit for determining in an individual lung cancer or for monitoring the progress of a lung cancer treatment in an individual, comprising reagent for detecting differential expression of at least 2 of the biomarker genes FEN1, RIPK2 and CCNE1, wherein one of the genes is FEN1 , optionally in combination with at least 1 and up to 108 different genes depicted in table 1, table 2, or both table 1 and table 2.
Methods, compositions and kits as defined above, wherein the biomarker genes comprise all three of FEN1, RIPK2, CCNE1.

---

2. claims: 1-9, 11-15, 17-21, 23-28, 30, 31, 33-35(all partially)

Methods of providing a prognosis on the development of lung cancer in an individual, and methods for predicting or monitoring the outcome/progress of a lung cancer treatment in an individual, said methods comprising determining in a sample from an individual a test value of the expression of a panel of biomarkers corresponding to at least 2 and up to 111 different genes depicted in table 1, table 2, or both table 1 and table 2, wherein at least one of the biomarker genes is RIPK2 , or wherein the markers compriseRIPK2 and one or two selected fromFEN1 andCCNE1 ,and comparing the values obtained for each biomarker with reference values obtained from lung cancer of a non-smoker and/or a smoker.
Compositions comprising comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide that corresponds to at least 2 of the biomarker genes FEN1, RIPK2 and CCNE1, wherein one of the genes is RIPK2 , and optionally in combination with at

least 1 and up to 108 different biomarker genes depicted in table 1, table 2, or both table 1 and table 2.
A kit for determining in an individual lung cancer or for monitoring the progress of a lung cancer treatment in an individual, comprising reagent for detecting differential expression of at least 2 of the biomarker genes FEN1, RIPK2 and CCNE1, wherein one of the genes is RIPK2 , optionally in combination with at least 1 and up to 108 different genes depicted in table 1, table 2, or both table 1 and table 2.

---

3. claims: 1-9, 11-15, 17-21, 23-28, 30, 31, 33-35(all partially)

Methods of providing a prognosis on the development of lung cancer in an individual, and methods for predicting or monitoring the outcome/progress of a lung cancer treatment in an individual, said methods comprising determining in a sample from an individual a test value of the expression of a panel of biomarkers corresponding to at least 2 and up to 111 different genes depicted in table 1, table 2, or both table 1 and table 2, wherein at least one of the biomarker genes is CCNE1 , or wherein the markers compriseCCNE1 and one or two selected fromFEN1 andRIPK2 ,and comparing the values obtained for each biomarker with reference values obtained from lung cancer of a non-smoker and/or a smoker.
Compositions comprising comprising at least 2 different isolated nucleic acid molecules each of which comprising a biomarker polynucleotide that corresponds to at least 2 of the biomarker genes FEN1, RIPK2 and CCNE1, wherein one of the genes is CCNE1 , and optionally in combination with at least 1 and up to 108 different biomarker genes depicted in table 1, table 2, or both table 1 and table 2.
A kit for determining in an individual lung cancer or for monitoring the progress of a lung cancer treatment in an individual, comprising reagent for detecting differential expression of at least 2 of the biomarker genes FEN1, RIPK2 and CCNE1, wherein one of the genes is CCNE1 , optionally in combination with at least 1 and up to 108 different genes depicted in table 1, table 2, or both table 1 and table 2.

4-111. claims: 33-35(partially)

Method of providing a prognosis on the development of lung cancer in an individual, methods for predicting or monitoring the outcome/progress of a lung cancer treatment in an individual, said methods comprising determining in a sample from a lung cancer patient a test value of the expression of each of at least 2 and up to 111 different biomarkers depicted in table 1, table 2, or both table 1 and table 2, wherein at least one of the biomarker genes is the first (second, third, fourth, ... 111th gene of tables 1 and 2 except for the genes FEN1, RIPK2, CCNE1) and comparing the values obtained for each biomarker with reference values obtained from lung cancer of a non-smoker.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2013/060809

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2010099161 A1	02-09-2010	AU 2010218147 A1	20-10-2011
		CA 2753562 A1	02-09-2010
		CN 102549166 A	04-07-2012
		EP 2401405 A1	04-01-2012
		JP 2012518997 A	23-08-2012
		US 2012027753 A1	02-02-2012
		WO 2010099161 A1	02-09-2010