



(51) International Patent Classification:

C12Q 1/68 (2006.01) C40B 40/06 (2006.01)
C40B 30/04 (2006.01) G01N 33/48 (2006.01)

(21) International Application Number:

PCT/CA2013/000822

(22) International Filing Date:

26 September 2013 (26.09.2013)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/706,261 27 September 2012 (27.09.2012) US

(71) Applicant: UNIVERSITE LAVAL [CA/CA]; Local
1434, 2320 rue des Bibliothèques, Quebec, Québec G1V
0A6 (CA).

(72) Inventors; and

(71) Applicants (for US only): BOSSE, Yohan [CA/CA];
9455 rue de Salzbouurg, Quebec, Québec G2B 0B6 (CA).
LAVIOLETTE, Michel [CA/CA]; 1085 rue de St-Sebas-
tien, Quebec, Québec G1V 4G5 (CA). JOUBERT, Phil-
ippe [CA/CA]; 20 Montée du St-Castin, Lac-Beauport,
québec G3B 0Z4 (CA). CARRIER, Jean-Sebastien
[CA/CA]; 16405 du Jade, Mirabel, Québec J7N 0J7 (CA).

(74) Agent: NORTON ROSE FULBRIGHT S.E.N.C.R.L.,
S.R.L.; Bureau 2500, 1, Place Ville-Marie, Montreal,
Québec H3B 1R1 (CA).

(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, 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, SA,
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,
KM, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: GENE EXPRESSION PROFILING FOR PROGNOSIS OF NON-SMALL CELL LUNG CANCER

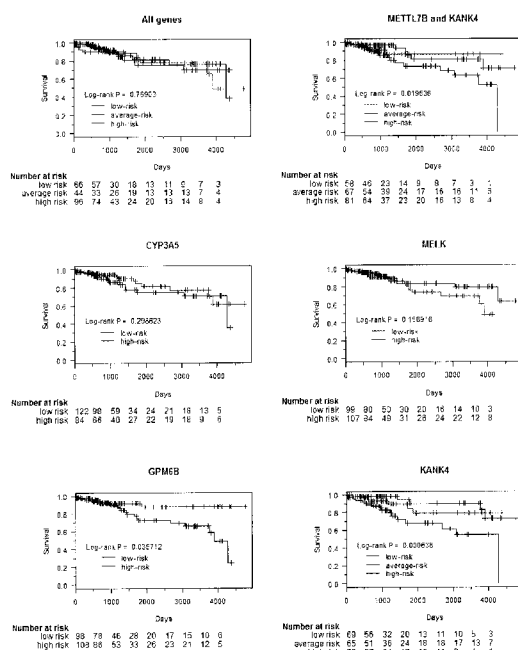


Figure 2

(57) Abstract: Twenty-two (22) genes associ-
ated with survival were identified using uni-
variate Cox proportional hazards regression
analyses (false discovery rate < 1%). Signifi-
cant genes were then validated by quantitative
real-time PCR (qPCR) in the same lung spe-
cimens in order to develop a qPCR assay with
optimal discriminatory performance. Six risk
models were developed from qPCR-based
gene expression levels which allowed delin-
eation of a high-risk group following stage I
adenocarcinoma resection (Kaplan-Meier log-
rank $p < 0.005$). The lock down qPCR-based
models were then validated in a new set of
206 non-tumor lung specimens explanted
from patients with confirmed stage I adeno-
carcinoma. The best risk model was based on
the mRNA expression level of a single gene,
namely KN motif and ankyrin repeat domains
4 (KANK4) (Kaplan-Meier log-rank $p =$
0.0006) and outperformed predictor of sur-
vival based on current pathologic staging
(Wilcoxon rank sum test $p < 1E-5$). This new
qPCR assay is needed in clinical setting to
identify high-risk patients following stage I
adenocarcinoma resection that will benefit

from adjuvant chemotherapy.



Published:

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

GENE EXPRESSION PROFILING FOR PROGNOSIS OF NON-SMALL CELL LUNG CANCER

Field of the invention

5 **[0001]** The present invention relates to markers for the prognosis of lung cancer progression or recurrence and adjustment of therapy. Particularly, the present invention relates to the measurement of KANK4 mRNA expression level in non-tumor lung tissues to predict survival of patients with non-small cell lung cancer (NSCLC), particularly stage I lung adenocarcinoma.

Background of the invention

10 **[0002]** Lung cancer is the leading cause of cancer deaths worldwide, with more than a million new cases annually¹. The histopathology TNM (Tumor-Node-Metastasis) staging system is currently used to guide treatment decisions for patients with non-small cell lung cancers (NSCLC)². For a technically resectable lesion, surgery remains the first line of treatment. A significant
15 minority of patients (25-30%) with early-stages NSCLC receive surgical intervention alone. However 35-50% of these patients will relapse within 5 years³, which suggests that the TNM staging system is insufficient to guide adjuvant chemotherapy and additional prognostic factors are urgently needed beyond existing diagnostic standard of care.

20 **[0003]** Gene expression profiling offers great promise to refine the TNM staging system and improve therapeutic decisions following lung resection⁴. Gene expression profiling studies so far have focused on prognostic gene expression signatures in surgically excised tumor samples with a range of histologies and stages⁴⁻²⁴. Tumors are heterogeneous tissues containing
25 multiple subclones and varying levels of cancer and normal cells²⁵, characteristics that make the development and validation of a robust gene expression signature challenging. Other studies have shown that lung cancer signatures can be detected in the entire respiratory track, including airway epithelial cells²⁶. Accordingly, we hypothesized that non-tumor lung
30 parenchyma derived from patients undergoing lung cancer surgery bears gene expression signatures that can be used to predict relapse-free and overall

survival. In this document we focus particularly on stage I adenocarcinoma. This subtype and stage of lung cancer represent the largest proportion of patients undergoing lung resection and will be increasingly detected with improved molecular and imaging strategies as well as more aggressive
5 computed tomography screening for high-risk patients²⁷.

Summary of the invention

[0004] In a first aspect of the invention, there is provided a method for prognosing a cancer recurrence in a human subject having undergone surgical resection of a non-small cell lung cancer; the method comprising:

- 10 a) determining **KANK4** gene (entrez # 163782) expression level present in a non-tumor lung sample derived from the subject; and
- b) comparing said level to a control level of **KANK4** gene present in a control sample;
- wherein a decreased level of said **KANK4** expression relative to the
15 level in the control sample indicates that the non-small cell lung cancer is at high risk of recurring in said subject.

[0005] In an alternative aspect of the invention, there is provided a method for prognosing a cancer recurrence in a human subject having undergone surgical resection of a non-small cell lung cancer; the method comprising:

- 20 a) determining a gene expression profile present in a non-tumor lung sample derived from the subject, the gene expression profile comprising the level of expression of at least one gene selected from the group consisting of: **ABHD9** (also 3BHD9 or EPHX3): epoxide hydrolase 3; **CLEC3B**: C-type lectin domain family 3, member B; **HIST3H2A**: histone
25 cluster 3, H2a; **LRP11**: low density lipoprotein receptor-related protein 11; **METTL7B**: methyltransferase like 7B; **ARHGEF6**: Rac/Cdc42 guanine nucleotide exchange factor (GEF) 6; **ATP2A2**: ATPase, Ca⁺⁺ transporting, cardiac muscle, slow twitch 2; **GPM6B**: glycoprotein M6B; **PIK3R1**: phosphoinositide-3-kinase, regulatory subunit 1 (alpha);
30 **RXFP1**: relaxin/insulin-like family peptide receptor 1; **CYP3A5**:

cytochrome P450, family 3, subfamily A, polypeptide 5; **KIT**: v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog; **PKIA**: protein kinase (cAMP-dependent, catalytic) inhibitor alpha; **KANK4**: KN motif and ankyrin repeat domains 4 (entrez # 163782); **P2RY14**: purinergic receptor P2Y, G-protein coupled, 14; **IGSF9**: immunoglobulin superfamily, member 9; **ATP6V1C2**: ATPase, H⁺ transporting, lysosomal 42kDa, V1 subunit C2; and **MELK**: maternal embryonic leucine zipper kinase; and

b) comparing said level to a control level of same said gene present in a control sample;

wherein a decreased or increased level of said gene expression relative to the level in the control sample indicates the risk of recurrence of said non-small cell lung cancer in said subject.

[0006] Particularly, the gene expression profile comprises at least one or a combination of genes such as for example, at least two, at least 3, at least 5, at least 10 genes. More particularly, the gene expression profile comprises: **KANK4** and/or **GPM6B** and/or **CYP3A5** and/or **MELK** and/or **METTL7B**. Most particularly, the gene expression profiles comprises: **KANK4**; or **GPM6B**; or **CYP3A5**; or **MELK**; or a combination of **METTL7B** and **KANK4**; or a combination of all 18 genes herein defined.

[0007] Particularly, the gene expression level consist of **GPM6B**; **CYP3A5**; **MELK**; a combination of **METTL7B** and **KANK4**; or a combination of all 18 genes listed herein above.

[0008] In a further aspect, the invention provides the method as defined above further comprising classifying said subject, whereby a difference in the expression profile of the gene in said sample is used for classifying the subject with NSCLC into a poor survival group (high-risk), an average risk or a good survival group (low risk).

[0009] The present invention also provides a composition comprising a plurality of isolated nucleic acid sequences, wherein each isolated nucleic acid sequence hybridizes to:

- a) at least one RNA product of one of the genes as defined herein;
and/or
- b) a nucleic acid complementary to a),

wherein the composition is used to measure the level of RNA expression of at

5 least one of said genes.

[0010] In a further aspect, the invention provides a kit for prognosing or classifying a subject with non-small cell lung carcinoma (NSCLC) comprising a set of detection agents capable of measuring the expression level of at least one gene in a test sample, wherein the gene is selected from the group

10 consisting of: **ABHD9** (also 3BHD9 or EPHX3): epoxide hydrolase 3; **CLEC3B**: C-type lectin domain family 3, member B; **HIST3H2A**: histone cluster 3, H2a; **LRP11**: low density lipoprotein receptor-related protein 11; **METTL7B**: methyltransferase like 7B; **ARHGEF6**: Rac/Cdc42 guanine nucleotide exchange factor (GEF) 6; **ATP2A2**: ATPase, Ca⁺⁺ transporting, cardiac

15 muscle, slow twitch 2; **GPM6B**: glycoprotein M6B; **PIK3R1**: phosphoinositide-3-kinase, regulatory subunit 1 (alpha); **RXFP1**: relaxin/insulin-like family peptide receptor 1; **CYP3A5**: cytochrome P450, family 3, subfamily A, polypeptide 5; **KIT**: v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog; **PKIA**: protein kinase (cAMP-dependent, catalytic) inhibitor alpha; **KANK4**: KN motif

20 and ankyrin repeat domains 4 (entrez # 163782); **P2RY14**: purinergic receptor P2Y, G-protein coupled, 14; **IGSF9**: immunoglobulin superfamily, member 9; **ATP6V1C2**: ATPase, H⁺ transporting, lysosomal 42kDa, V1 subunit C2; and **MELK**: maternal embryonic leucine zipper kinase.

[0011] Particularly, in accordance with the kit of the invention, the

25 detection agent is capable of measuring the expression level of: at least one, at least two, at least 3, at least 5, at least 10 genes from **Table 8**. Particularly, the gene expression profile comprises: **KANK4** and/or **GPM6B** and/or **CYP3A5** and/or **MELK** and/or **METTL7B**. More particularly, the gene expression

30 profiles comprises: **KANK4**; or **GPM6B**; or **CYP3A5**; or **MELK**; or a combination of **METTL7B** and **KANK4**; or a combination of all 18 genes herein defined.

[0012] In a further aspect, the invention provides a method for treating a human subject being prognosed for cancer recurrence of a non-small cell lung cancer, the method comprising:

5 a) determining a gene expression profile present in a non-tumor lung sample derived from the subject, the gene expression profile comprising the level of expression of at least one gene selected from the group consisting of: **ABHD9** (also 3BHD9 or EPHX3): epoxide hydrolase 3; **CLEC3B**: C-type lectin domain family 3, member B; **HIST3H2A**: histone cluster 3, H2a; **LRP11**: low density lipoprotein receptor-related protein
10 11; **METTL7B**: methyltransferase like 7B; **ARHGEF6**: Rac/Cdc42 guanine nucleotide exchange factor (GEF) 6; **ATP2A2**: ATPase, Ca⁺⁺ transporting, cardiac muscle, slow twitch 2; **GPM6B**: glycoprotein M6B; **PIK3R1**: phosphoinositide-3-kinase, regulatory subunit 1 (alpha); **RXFP1**: relaxin/insulin-like family peptide receptor 1; **CYP3A5**:
15 cytochrome P450, family 3, subfamily A, polypeptide 5; **KIT**: v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog; **PKIA**: protein kinase (cAMP-dependent, catalytic) inhibitor alpha; **KANK4**: KN motif and ankyrin repeat domains 4 (entrez # 163782); **P2RY14**: purinergic receptor P2Y, G-protein coupled, 14; **IGSF9**: immunoglobulin
20 superfamily, member 9; **ATP6V1C2**: ATPase, H⁺ transporting, lysosomal 42kDa, V1 subunit C2; and **MELK**: maternal embryonic leucine zipper kinase;

and

25 b) comparing said level to a control level of same said gene present in a control sample;

wherein a decreased or increased level of said gene expression relative to the level in the control sample indicates that the non-small cell lung cancer is at high-risk of recurring in said subject, whereby the prognosed cancer recurrence in the subject will be treated.

30 **[0013]** Particularly, in accordance with the method of treatment, the gene expression profile comprises at least one or a combination of genes such as for

example, at least two, at least 3, at least 5, at least 10 genes. Particularly, the gene expression profile comprises: **KANK4** and/or **GPM6B** and/or **CYP3A5** and/or **MELK** and/or **METTL7B**. More particularly, the gene expression profiles comprises: **KANK4**; or **GPM6B**; or **CYP3A5**; or **MELK**; or a combination of **METTL7B** and **KANK4**; or a combination of all 18 genes herein defined.

[0014] According to an alternative aspect of the present invention, there is provided a biomarker for predicting the likelihood of non-small cell lung cancer (NSCLC) progression or recurrence in a human subject, comprising the identification of a variation in expression level of a gene in normal lung tissue from said subject, the gene being selected from the group consisting of: P2RY14; GPM6B; KANK4; IGSF9; MELK; HIST3H2A; PKIA; KIT; METTL7B; ATP2A2; ARHGEF6; RXFP1; PIK3R1; ATP6V1C2; CLEC3B; ABHD9; CYP3A5; and LRP11.

[0015] According to a further aspect of the present invention, there is provided a plurality of biomarkers for predicting survival of a subject with non-small cell lung cancer (NSCLC), the plurality of biomarkers comprising at least one of: P2RY14; GPM6B; KANK4; IGSF9; MELK; HIST3H2A; PKIA; KIT; METTL7B; ATP2A2; ARHGEF6; RXFP1; PIK3R1; ATP6V1C2; CLEC3B; ABHD9; CYP3A5; and LRP11.

[0016] According to a further aspect of the present invention, there is provided a method for predicting survival of a patient with non-small cell lung cancer (NSCLC), the method comprising: a) determining a gene expressing profile from a sample of the patient's non-tumor lung, the gene expression profile comprising the level of expression of at least one gene from **Table 8**, and b) classifying the gene expression profile as being predictive of good survival (low-risk), average risk or poor survival (high-risk).

[0017] According to a further aspect of the present invention, there is provided a method for preparing a gene expression profile indicative of the need for adjuvant chemotherapy for non-small cell lung cancer (NSCLC) patient, comprising: a) determining the level of expression of at least one gene

from **Table 8** from a non-tumor lung tissue sample from said patient with NSCLC.

[0018] According to a further aspect of the present invention, there is provided a method of prognosing or classifying a subject with non-small cell lung cancer (NSCLC) comprising the steps: a) determining the expression of a biomarker in a test sample from the subject, wherein the biomarker correspond to at least one gene in **Table 8**, and b) comparing the expression of the biomarkers in the test sample with expression of reference genes in said sample, wherein a difference in the expression of the biomarkers in samples is used for prognosis or classify the subject with NSCLC into a poor survival group (high-risk) or a good survival group (low risk).

[0019] According to a further aspect of the present invention, there is provided a method of prognosing or classifying a subject with non-small cell lung cancer comprising: a) determining the expression of at least one biomarker in a test sample from the subject, wherein the biomarkers correspond to genes in **Table 8**, b) combining the expression of each of the biomarkers by a regression coefficient for the corresponding biomarker, c) calculating a risk score for the test sample by summing the values obtained in step (b), and d) deriving the low and high risk groups by comparing the risk score to the risk category cut-off value, wherein a risk score cut-off value is used for prognosis or classify the subject with NSCLC into a poor or good survival groups.

[0020] According to a further aspect of the present invention, there is provided a method of prognosing a subject with NSCLC comprising: a) determining expression levels of at least one biomarkers from Table 8, b) calculating a risk score for the subject from the expression levels of said biomarkers, and c) comparing the risk score to a cut-off value, wherein a risk score cut-off value is used to classify a subject into a high risk (poor survival) or low risk (good survival) groups.

[0021] According to a further aspect of the present invention, there is provided a composition comprising a plurality of isolated nucleic acid sequences, wherein each isolated nucleic acid sequence hybridizes to: a) an RNA product of one of the 18 genes listed in Table 8; and/or b) a nucleic acid

complementary to a), wherein the composition is used to measure the level of RNA expression of each of the 18 the gene listed in Table 8, or a subset thereof.

5 **[0022]** According to a further aspect of the present invention, there is provided an array comprising at least one polynucleotide probe hybridizable to an expression product of each of the 18 the genes listed in Table 8, or a subset thereof.

10 **[0023]** According to a further aspect of the present invention, there is provided a computer program product for use in conjunction with a computer having a processor and a memory connected to the processor, the computer program product comprising a computer readable storage medium having a computer mechanism encoded thereon, wherein the computer program mechanism may be loaded into the memory of the computer and cause the computer to carry out the method as defined herein.

15 **[0024]** According to a further aspect of the present invention, there is provided a computer product for implementing a prognostic algorithm predicting a prognosis or classifying a subject with NSCLC comprising: a) a means for receiving values corresponding to a subject expression profile in a subject sample; and b) a database comprising a complete set of algorithm coefficients and cut-off values associated with a prognosis, wherein each value
20 representing the expression level of a biomarker, wherein each biomarker corresponds to one gene in Table 8; wherein the computer product selects the risk group category to the subject based on biomarker expression profile, thereby predicting a prognosis or classifying the subject.

25 **[0025]** According to a further aspect of the present invention, there is provided a computer readable medium having stored thereon a data structure for storing the computer implemented product as defined above.

30 **[0026]** According to a further aspect of the present invention, there is provided a computer system comprising: a) a complete set of algorithm coefficients and cut-off values to derive the risk categories based on expression values of 18 genes in Table 8; b) a user interface capable of receiving a

selection of gene expression levels of the 18 genes in Table 8 to calculate the risk score and discriminate patients into a low and high risk groups, and c) an output that displays a prediction of prognosis or therapy according to the biomarker expression profile of the 18 genes.

- 5 **[0027]** According to a further aspect of the present invention, there is provided a kit for prognosis or classify a subject with non-small cell lung carcinoma (NSCLC) comprising a set of detection agents capable of detecting the expression products of at least one biomarker in a test sample, wherein the biomarkers are selected from the group consisting of: P2RY14; GPM6B;
10 KANK4; IGSF9; MELK; HIST3H2A; PKIA; KIT; METTL7B; ATP2A2; ARHGEF6; RXFP1; PIK3R1; ATP6V1C2; CLEC3B; ABHD9; CYP3A5; and LRP11.

- [0028]** According to a further aspect of the present invention, there is provided a method for adapting a course of treatment of lung cancer in a human subject, comprising the steps of: a) predicting the likelihood of lung
15 cancer recurrence in a human subject in accordance with the method as defined herein, and b) adapting a course of treatment according to whether said subject is at low-risk or high-risk of recurrence, and whereby said subject at high risk will be further treated.

Detailed description of the invention

20 ***Description of the figures***

[0029] **Figure 1.** Kaplan-Meier analysis for the six most discriminatory risk categories in the discovery set 2. Titles represent genes used to calculate the genetic risk score.

- [0030]** **Figure 2.** Kaplan-Meier analysis of overall survival in the validation
25 set according to the six most discriminatory risk categories identified in the validation set. A total of 206 patients with stage I adenocarcinoma are considered in this analysis.

Abbreviations and Definitions

Abbreviations

[0031] **P2RY14**: purinergic receptor P2Y, G-protein coupled, 14; **GPM6B**: glycoprotein M6B; **KANK4**: KN motif and ankyrin repeat domains 4 (entrez # 163782); **IGSF9**: immunoglobulin superfamily, member 9; **MELK**: maternal embryonic leucine zipper kinase; **UBE2C**: ubiquitin-conjugated enzyme E2C; **HIST3H2A**: histone cluster 3, H2a; **PKIA**: protein kinase (cAMP-dependent, catalytic) inhibitor alpha; **KIT**: v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog; **DNAJC22**: DnaJ (Hsp40) homolog, subfamily C, member 22; **METTL7B**: methyltransferase like 7B; **ATP2A2**: ATPase, Ca⁺⁺ transporting, cardiac muscle, slow twitch 2; **ARHGEF6**: Rac/Cdc42 guanine nucleotide exchange factor (GEF) 6; **RXFP1**: relaxin/insulin-like family peptide receptor 1; **PIK3R1**: phosphoinositide-3-kinase, regulatory subunit 1 (alpha); **MF12**: antigen p97 (melanoma associated) identified by monoclonal antibodies 133.2 and 96.5; **ATP6V1C2**: ATPase, H⁺ transporting, lysosomal 42kDa, V1 subunit C2; **CLEC3B**: C-type lectin domain family 3, member B; **ABHD9** (also 3BHD9 or EPHX3): epoxide hydrolase 3; **CYP3A5**: cytochrome P450, family 3, subfamily A, polypeptide 5; **REXO1L1** (also REX1): RNA exonuclease 1 homolog (*S. cerevisiae*)-like 1; **LRP11**: low density lipoprotein receptor-related protein 11.

Detailed description of particular embodiments

Method

[0032] According to a first aspect of the present invention, there is provided a biomarker and a method for predicting the likelihood of lung cancer relapse and overall survival following lung cancer resection surgery in a human subject, comprising the identification in normal lung tissue from said human subject of a variation in a gene selected from the group consisting of: P2RY14; GPM6B; KANK4; IGSF9; MELK; HIST3H2A; PKIA; KIT; METTL7B; ATP2A2; ARHGEF6; RXFP1; PIK3R1; ATP6V1C2; CLEC3B; ABHD9; CYP3A5; and LRP11.

[0033] Particularly, there is provided a method for predicting survival of a patient with non-small cell lung cancer (NSCLC), the method comprising: a)

determining a gene expressing profile from a sample of the patient's non-tumor lung, the gene expression profile comprising the level of expression of at least one gene from the 18 genes defined in Table 8, and b) classifying the gene expression profile as being predictive of good survival or poor survival.

- 5 **[0034]** More particularly, the method comprises determining the level of expression of at least one gene from Table 8, still more particularly, at least 2 genes or at least 3 genes or at least 5 genes or at least 10 genes or at least 15 genes. Still, more particularly, the expression levels for all 18 genes from Table 8 are determined.
- 10 **[0035]** Particularly, the gene expression profile comprises or consists of: GPM6B.
- [0036]** Particularly, the gene expression profile comprises or consists of: CYP3A5.
- [0037]** Particularly, the gene expression profile comprises or consists of:
- 15 **MELK.**
- [0038]** Particularly, the gene expression profile comprises or consists of: a combination of METTL7B and KANK4.
- [0039]** Particularly, the gene expression profile comprises or consists of: a combination of all 18 genes listed herein above.
- 20 **[0040]** Most particularly, the gene expression profile comprises or consists of: the **KANK4** gene (entrez # 163782) .

Control level

- 25 **[0041]** In connection with the different aspects of the present invention, the comparing step b) is carried out once to establish a threshold, or every time an assay is carried out.
- [0042]** Particularly, the control level is obtained from a control sample, wherein the control level of the selected gene is derived from a cell or a sample (such as lung) from a subject not afflicted with non-small cell lung cancer; or is obtained from a cell or a sample (such as lung) from a subject afflicted with

non-small cell lung cancer but for which a low-risk or high survival has been established.

Reference genes

[0043] Particularly, there is provided the method as defined herein,
5 wherein the gene expression profile is normalized with the use of average expression level of one or more reference genes in the sample. More particularly, the reference gene is selected from the group consisting of: **TBP** (TATA box binding protein), **GAPDH** (glyceraldehyde-3-phosphate dehydrogenase), **HRPT1** (hyperparathyroidism 1), **GUSB** (beta-glucuronidase)
10 and **ACTB** (beta-actin).

Subject

[0044] Particularly, in accordance with the method of present invention, the subject is a human. More particularly, the subject has a stage I lung adenocarcinoma.

15 Biological sample

[0045] Particularly, there is provided the method as defined herein, wherein the test sample from the subject is a non-tumor sample. More particularly, the sample is a tissue or biological fluid, alternatively sputum, saliva or alveolar lavage. More particularly, the sample is a non-tumor lung tissue or a
20 non-neoplastic pulmonary parenchyma sample.

Kit

[0046] There is also provided a kit for performing the method as defined herein. In this kit, the method is qPCR carried out on nucleic acid, such as DNA or RNA with prior amplification. Particularly, the kit comprises reagents. More
25 particularly, the kit is useful for prognosis or classify a subject with non-small cell lung carcinoma (NSCLC) and comprises a set of detection agents capable of detecting the expression products of a set number of biomarkers in a test sample, wherein the biomarkers are selected from the group consisting of:
P2RY14; GPM6B; KANK4; IGSF9; MELK; HIST3H2A; PKIA; KIT; METTL7B;
30 ATP2A2; ARHGEF6; RXFP1; PIK3R1; ATP6V1C2; CLEC3B; ABHD9; CYP3A5; and LRP11.

[0047] More particularly, the detection agents are nucleic acid probes as defined herein, more particularly as defined in **Table 7**. More particularly, the reagents are PCR primer-probe sets, wherein the primer is selected from the group consisting of: SEQ ID Nos. 11 to 58.

5 Computer

[0048] The invention also provides for a computer program product for use in conjunction with a computer having a processor and a memory connected to the processor to carry out the method as defined herein. Particularly, there is provided a computer implemented product for predicting a prognosis or
10 classifying a subject with NSCLC, wherein the computer implemented product selects the biomarker reference expression profile most similar to the subject biomarker expression profile, to thereby predict a prognosis or classify the subject. Additionally, there is provided a computer readable medium having stored thereon a data structure for storing the computer implemented product
15 as defined above, where the data structure is capable of configuring a computer to respond to queries based on records belonging to the data structure, each of the records comprising: a) a value that identifies a biomarker reference expression profile of the 18 genes in Table 8; b) a value that identifies the probability of a prognosis associated with the biomarker reference expression
20 profile.

Adapting course of treatment

[0049] There is also provided a method for adapting a course of treatment of lung cancer in a human subject, comprising the steps of: a) predicting the likelihood of lung cancer progression or recurrence in a human subject in
25 accordance with the method as defined herein; and b) adapting a course of treatment according to whether said subject is a low risk, average risk or high-risk of recurrence. Particularly, when the subject is determined to be at high-risk, the subject is prescribed at least one of: chemotherapy, radiotherapy, active surveillance (watchful waiting). Most particularly, steps a) and b) are
30 carried out after the subject has undergone lung resection.

[0050] The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to

make and use the present invention, and are not intended to limit the scope of what the inventors regard as their invention nor are they intended to represent that the experiments below are all or the only experiments performed. Efforts have been made to ensure accuracy with respect to numbers used (e.g. amounts, temperature, etc.) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, molecular weight is weight average molecular weight, temperature is in degrees Centigrade, and pressure is at or near atmospheric.

Examples

10 **Example 1**

Materials and methods

Study participants

[0051] Frozen lung specimens for the discovery and validation sets were obtained from the Respiratory Health Network Tissue Bank of the FRQS (tissuebank.ca), site *Institut Universitaire de Cardiologie et de Pneumologie de Québec* (IUCPQ). Lung tissues for the discovery set were explanted from patients who underwent surgical resection at the IUCPQ between February 2004 and December 2008. Only patients with stage I adenocarcinoma were considered. Staging was defined based on the seventh edition of the TNM cancer staging manual²⁸. Similarly, the validation set consisted of patients with stage I adenocarcinoma resected at the IUCPQ between December 1999 and September 2011. Corresponding clinical variables including demographic data, pathology report and smoking status were taken from the biobank database. Patient's medical charts were abstracted for follow-up starting at the time of surgery, vital status, date and cause of death. Missing data from the file were obtained by contacting patients or their family. Patients were observed until death or last follow-up performed in July 2013. Exclusion criteria for the discovery and validation sets include never smokers, positive neoplastic margins on resected lung tissue and previous cancer of any origins with systemic treatment within 5 years of lung cancer surgery. A representative histologic slide of non-tumor tissue was reviewed by a pathologist (P.J.) and specimens with tumor cells were excluded from the analysis. Following these

quality control filters, 116 and 206 patients were included in the discovery and validation sets, respectively. **Table 1** shows the clinical characteristics of both cohorts used in survival analyses. All patients provided written informed consent, and the study was approved by the ethics committee of the IUCPQ.

5 *Lung specimen*

[0052] Non-tumor lung specimens explanted during surgery were immediately examined by an experienced pathologist for clinical diagnosis and staging. For each specimen, a non-neoplastic pulmonary parenchyma sample was harvested from a site as distant as possible from the tumor, snap-frozen in
10 liquid nitrogen, and stored at -80°C until further processing.

Genome-wide gene expression in the lung

[0053] Total RNA from specimens in the discovery set 1 was extracted using the SV96 Total RNA Isolation System (Promega). Expression profiling was performed using an Affymetrix custom array (GEO platform GPL10379)
15 and testing 51,562 probe sets. Expression values were extracted using the Robust Multichip Average (RMA) method²⁹ as implemented in the Affymetrix Power Tools (APT) software. The quality of the arrays was judged using standard quality control parameters.

Quantitative real-time PCR (qPCR)

20 **[0054]** RNA from specimens in the discovery set 2 and validation set was extracted from 30 mg of frozen lung parenchyma specimens using RNeasy Mini kit (Qiagen). RNA concentration and purity was verified by UV 260/280 nm ratio on the NanoVue spectrophotometer (GE Healthcare). Six micrograms of RNA were converted to cDNA using Quantitect Reverse Transcription kit (Qiagen).
25 qPCR was performed on the ABI Prism 7900HT system in 384-well plate format. Distribution of samples into the plates was performed using an 8-channel multipipette, and distribution of the reaction mix was performed using the epMotion 5075 robot (Eppendorf). The total reaction volume was 12 µl containing 6 µl of SYBR Green JumpStart Taq ReadyMix (Sigma-Aldrich), 0.48
30 µl of ROX Reference Dye for Quantitative PCR 100X solution (Sigma-Aldrich), 1.26 µl of 25 mM MgCl₂, 0.72 µl of a 5 µM mix of each primer, 0.54 µl of water and 3 µl of 50X diluted cDNA or as specified in **Table 7**. Cycling conditions

were as follows: 2 min denaturation at 94°C, and 40 amplification cycles consisting of 15 s at 94°C and 1 min at 60°C. Primers were designed using Primer3 ([//frodo.wi.mit.edu/primer3/](http://frodo.wi.mit.edu/primer3/)) and ordered from Integrated DNA Technologies. Five reference genes were considered including TBP, HPRT1, GAPDH, GUSB1 and ACTB. Primers for the targeted genes and five reference genes are shown in **Table 7**. Expression values were calibrated according to the standard curve. The standard curves were constructed with a minimum of seven points using serial dilutions (undiluted to 500x) from a pool of the 95 experimental samples. A total of 318 qPCR reactions were performed for each gene including 3 replicates of the following: the corresponding standard curve, a negative control containing no cDNA, a positive control and each of the experimental samples. The targeted acceptance criteria for the calibration curve were as follows: $r^2 \geq 0.99$ and efficacy between 90 and 110%. Samples were re-tested if the CV between replicates were >17% and/or found outside the upper and lower limits of the calibration curve. The average expression of the five reference genes was used to normalize expression of the targeted genes. qPCR methods were repeated for the biological validation for 206 experimental samples.

Statistical analyses

20 Discovery set 1

[0055] Our primary endpoint was overall survival from time of resection. Univariate Cox proportional hazards regression analyses were used to evaluate the association between survival and the level of expression of each gene evaluated on the microarray platform. The Benjamini-Hochberg procedure was used to calculate the false discovery rate (FDR)³¹. Well-annotated genes with FDR < 1% were considered significant. Significant genes associated with survival were then transferred into a qPCR assay using the same lung specimen (i.e. discovery set 1).

Discovery set 2 (technical validation)

30 **[0056]** Univariate and multivariate Cox proportional hazards regression were performed on pPCR-based expression values for technical validation. Multivariate models include genes either all present or selected based on the

following algorithms: backward selection, forward selection or penalized Cox regression. Continuous risk score for each patient were calculated from the gene expression values weighted by the regression coefficients derived from the univariate models and from linear combinations of weighted gene expression for the multivariate models. Risk score were scaled on a 1 to 100 interval. Patients were then classified as having a high-risk gene signature or a low-risk gene signature based on the median value of the risk score or as having a high-risk, an average-risk and a low-risk based on the 33rd and 67th tertile values of the risk score. Kaplan-Meier analysis was performed to evaluate the discriminatory performance of the risk categories on survival. Risk categories were then ranked according to log-rank test p-values. The six best performing categories were then evaluated in the validation cohort

Validation set.

[0057] Risk scores and risk categories were calculated for each participant based on the regression coefficients and the median or tertiles of the discovery set 2. Risk scores smaller than 1 or greater than 100 were replaced by 1 or 100, respectively. Discriminatory performance was evaluated by Kaplan-Meier analysis and log-rank test. The most significant risk categories (p-value < 0.01) were compared to pathologic staging: a Wilcoxon rank sum test was used to compare integrated area under the time-dependent receiver operator curve (iAUROC). Univariate and multivariate Cox proportional hazards regression models were used to assess the effect of available variables on overall survival. All analyses were performed with the R 2.15.3 statistical software³² using the following packages: survival³³, survcomp³⁴ and glmnet³⁵.

Example 2

Results

Discovery sets 1 and 2

The mean follow-up time of the discovery set 1 (n=116) for survival analysis was nearly 3 years (**Table 1**). The individual predicting value of all probe sets on the Affymetrix array (n = 51,562 probe sets) were evaluated using univariate Cox proportional hazards modelling. Twenty-two (22) genes had a FDR < 1%

and were deemed significant. The individual effect of these 22 genes on survival is showed in **Table 2**.

Table 1. Clinical characteristics of the discovery and validation cohorts

Characteristic	Discovery set 1	Discovery set 2	Validation set
No. of patients	116	95	206
Age (years; mean $\pm$ SD)	63.5 $\pm$ 8.6	63.6 $\pm$ 9.0	63.0 $\pm$ 9.1
Sex – no. of patients (%)			
Male	51 (44)	38 (40)	102 (50)
Female	65 (56)	57 (60)	104 (50)
Tumor stage – no. of patients (%)			
IA	65 (56)	53 (56)	101 (49)
IB	51 (44)	42 (44)	105 (51)
$\geq$ II	0 (0)	0 (0)	0 (0)
Tumor type – no. of patients (%)			
Adenocarcinoma	116 (100)	95 (100)	206 (100)
Other	0 (0)	0 (0)	0 (0)
Smoking status – no. of patients (%)			
Never-smokers	0 (0)	0 (0)	0 (0)
Former-smokers	88 (76)	72 (76)	138 (67)
Current-smokers	28 (24)	23 (24)	68 (33)
Follow-up (days; median, IQR)	1136 (771-1526)	1091 (769-1506)	970 (597 - 1788)
Deaths – no. of patients (%)	15 (13)	11 (12)	29 (14)

Table 2. Genes predicting survival at a false discovery rate (FDR) below 1% in the training set

Entrez #	Gene Symbol	p value*	Hazard ratio [95% CI]	Coef	FDR
9934	P2RY14	7.69E-008	0.026 [0.007, 0.099]	-3.664	0.0023
2824	GPM6B	8.90E-008	0.043 [0.014, 0.132]	-3.148	0.0023
163782	KANK4	8.55E-007	0.085 [0.032, 0.225]	-2.466	0.0047
57549	IGSF9	6.80E-007	81.441 [13.33, 497.566]	4.400	0.0047
9833	MELK	8.93E-007	7.479 [3.081, 18.155]	2.012	0.0047
11065	UBE2C	9.18E-007	6.927 [3.004, 15.973]	1.935	0.0047
92815	HIST3H2A	7.39E-007	43.367 [6.806, 276.317]	3.770	0.0047
5569	PKIA	4.23E-007	0.032 [0.009, 0.12]	-3.437	0.0047
3815	KIT	1.45E-006	0.066 [0.02, 0.223]	-2.714	0.0054
79962	DNAJC22	1.46E-006	24.754 [5.076, 120.724]	3.209	0.0054
196410	METTL7B	1.66E-006	5.428 [2.64, 11.16]	1.692	0.0057
488	ATP2A2	2.88E-006	135.108 [15.956, 1144.05]	4.906	0.0082
9459	ARHGEF6	2.68E-006	0.003 [0, 0.032]	-5.751	0.0082
59350	RXFP1	2.82E-006	0.139 [0.059, 0.328]	-1.975	0.0082
5295	PIK3R1	3.72E-006	0.033 [0.007, 0.151]	-3.415	0.0094
4241	MFI2	3.84E-006	10.833 [3.276, 35.823]	2.383	0.0094
245973	ATP6V1C2	4.03E-006	23.281 [5.976, 90.697]	3.148	0.0094
7123	CLEC3B	4.62E-006	0.072 [0.022, 0.233]	-2.630	0.0096
79852	ABHD9	4.74E-006	68.415 [9.233, 506.929]	4.226	0.0096
1577	CYP3A5	4.82E-006	0.152 [0.067, 0.345]	-1.886	0.0096
254958	REXO1L1	5.11E-006	0 [0, 0.009]	-9.808	0.0098
84918	LRP11	5.32E-006	40.972 [7.743, 216.814]	3.713	0.0098

*Likelihood ratio tests

Example 3qPCR assay

[0058] In order to develop a more practical and convenient prognostic gene expression signature, the 22 genes predicting survival on microarray-based expression values were validated by qPCR on lung specimens from the discovery set 1. 95 samples (**Table 1**) and 18 genes were available for the technical validation (discovery set 2), i.e. 19 samples were used up and 2 failed RNA quality control, four genes has unacceptable or unobtainable standard curves, nonspecific fragments amplified or instability between replicates. **Table 3** shows the results of the univariate Cox regressions, where 12 out of 18 genes were still associated with survival (5% threshold). **Table 4** shows the results of the multivariate Cox regressions using all genes or genes selected based on backward, forward and penalized Cox regression. Continuous risk scores were calculated from a linear combination of the qPCR-based gene expression values weighted by the regression coefficients and scaled according to minimum and maximum values presented in **Table 5**. Risk categories were obtained from median and tertile values presented in **Table 5**. Kaplan-Meier analyses of overall survival according to the risk categories were ranked according to log-rank test p-values (**Table 8**), and the six risk categories with the best discriminatory performance are presented in **Figure 1**. All six categories were strong predictors of overall survival with p-values bellow 0.005. Low risk groups had only one death or less and high risk groups had up to ten deaths in less than four years.

Table 3. Univariate regression coefficients from qPCR-based gene expression in the discovery set 2

Gene	Regression		P-value*
	coefficient	Hazard ratio [95% CI]	
KANK4	-3.3	0.0368 [0.00423, 0.32]	0.0004519
KIT	-3.15	0.043 [0.00402, 0.459]	0.001316
RXFP1	-2.67	0.0695 [0.00989, 0.488]	0.002821
GPM6B	-2.82	0.0594 [0.00667, 0.529]	0.002975
METTL7B	1.88	6.59 [2.15, 20.2]	0.00344
ARHGEF6	-2.59	0.0751 [0.0103, 0.547]	0.006183
PKIA	-3.27	0.0381 [0.00315, 0.46]	0.007339
ATP6V1C2	0.479	1.61 [1.17, 2.23]	0.0138
CYP3A5	-1.87	0.155 [0.0251, 0.953]	0.0141
CLEC3B	-2.05	0.128 [0.0183, 0.899]	0.01564
PIK3R1	-2.51	0.0812 [0.00745, 0.884]	0.01787
P2RY14	-2.13	0.119 [0.0145, 0.979]	0.03061
MELK	0.982	2.67 [1.04, 6.88]	0.05414
ATP2A2	1.86	6.42 [0.688, 59.9]	0.1305
ABHD9	1.34	3.84 [0.748, 19.7]	0.1483
HIST3H2A	0.624	1.87 [0.822, 4.24]	0.1911
IGSF9	0.278	1.32 [0.93, 1.87]	0.1915
LRP11	0.799	2.22 [0.322, 15.4]	0.4364

*Likelihood ratio tests

Table 4. Multivariate regression coefficients from qPCR-based gene expression in the discovery set 2

Model	Gene	Regression		
		coefficient	Hazard ratio [95% CI]	P-value
Backward	CLEC3B	-1.73	0.178 [0.023, 1.37]	0.09771
	METTL7B	1.45	4.26 [1.37, 13.2]	0.01222
	KIT	-2.62	0.0725 [0.00622, 0.846]	0.03629
Forward	METTL7B	1.43	4.17 [1.26, 13.8]	0.01968
	KANK4	-2.89	0.0557 [0.00579, 0.536]	0.0124
Penalized	ABHD9	0.748		
	GPM6B	-1.784		
All genes	ABHD9	3.64	38.2 [0.542, 2690]	0.0934
	ATP6V1C2	0.233	1.26 [0.682, 2.34]	0.4585
	CLEC3B	-1.65	0.193 [0.00874, 4.25]	0.2968
	HIST3H2A	-1.92	0.146 [0.0152, 1.41]	0.09648
	LRP11	-0.392	0.676 [0.00264, 173]	0.8899
	METTL7B	0.766	2.15 [0.433, 10.7]	0.3493
	ARHGEF6	-1.58	0.205 [0.00127, 33.1]	0.5415
	ATP2A2	-1.75	0.174 [0.000136, 221]	0.6313
	GPM6B	1.15	3.16 [0.0289, 346]	0.6307
	PIK3R1	1.79	5.96 [0.0941, 377]	0.399
	RXFP1	-1.78	0.168 [0.00266, 10.6]	0.3995
	CYP3A5	-1.29	0.275 [0.0306, 2.46]	0.248
	KIT	-2.88	0.0563 [0.00167, 1.9]	0.1091
	MELK	0.106	1.11 [0.141, 8.75]	0.9195
	PKIA	-2.36	0.0941 [0.000591, 15]	0.3608
	KANK4	-1.5	0.222 [0.00536, 9.22]	0.4288
	P2RY14	2.43	11.4 [0.241, 534]	0.2164
	IGSF9	0.0118	1.01 [0.374, 2.74]	0.9815

Table 5. Risk scores and cutoff values to define risk categories into median and tertile groups

Model	Raw risk score		Scaled risk score		
	Minimum	Maximum	Q33	Q50	Q67
ABHD9	0.294	3.057	21.353	25.601	30.894
CLEC3B	-6.907	-0.358	72.292	78.119	83.765
HIST3H2A	0.140	2.688	12.156	14.800	20.438
LRP11	0.497	1.672	21.474	27.112	35.518
METTL7B	0.553	5.162	15.721	19.988	23.157
ARHGEF6	-7.258	-0.536	61.509	67.985	73.931
ATP2A2	1.340	3.693	19.476	26.499	31.257
GPM6B	-8.003	-0.856	66.677	73.085	81.058
PIK3R1	-6.193	-0.989	63.754	70.707	78.993
RXFP1	-6.658	-0.346	53.770	62.190	67.350
CYP3A5	-4.118	-0.275	54.952	72.402	80.143
KIT	-7.228	-0.823	59.296	68.178	79.471
PKIA	-4.927	-0.936	42.279	49.795	60.544
KANK4	-7.759	-0.317	59.034	65.560	74.355
P2RY14	-4.068	-0.571	49.222	59.622	69.819
IGSF9	0.065	1.815	7.438	10.649	13.189
ATP6V1C2	0.115	2.956	9.559	12.751	18.176
MELK	0.125	2.782	21.125	28.174	37.394
All genes	-13.563	-1.283	41.253	52.443	59.129
Backward	-6.878	2.177	34.969	45.301	52.368
Forward	-5.906	3.055	43.701	50.091	58.776
Penalized	-4.203	1.164	51.025	57.885	62.938

Table 6. Cox proportional hazards modeling to predict overall survival in the validation set

	Univariate		Mutlivariate	
	Hazard Ratio	P-value*	Hazard ratio	P-value*
Risk Category based on <i>KANK4</i>				
Average risk	0.5516	0.3143	0.5531	0.0041
High risk	2.5012	0.0415	2.3514	
Age (>65)	2.2479	0.0329	2.1687	0.0480
Sex (male)	1.4880	0.2923	1.2556	0.5710
Smoking (smoker)	1.2123	0.6134	1.0549	0.7901
Staging (IB)	1.5912	0.2221	1.4855	0.2981

*Wald test for univariate models and maximum likelihood test for multivariate models

Table 7. Primers used for the quantitative real-time PCR

Gene	Forward oligo 5'-3'	SEQ ID NO.	Reverse oligo 5'-3'	SEQ ID NO.	Final oligo conc (μM)	Final MgCl ₂ conc (mM)	cDNA dilution factor
ACTB	ACTTCGAGCAAGAGATGGCCAC	1	GGACTCCATGCCAGGAAGG	2	0.3	3	20
GAPDH	ATGTCGTCATGGGTGTGAA	3	GGTGTAAGCAGTTGGTGT	4	0.3	2.625	50
GUSB	ACCAGCGTGGAGCAAGACAG	5	AAGTTGGCCCTGGGTCCC	6	0.3	2.625	50
HRPT1	ACCAGTCAACAGGGGACATAAA	7	CGACCTTGACCATCTTTGGATT	8	0.3	2.625	50
TBP	AGAACAAACAGCCTGCCACCTTAC	9	GTGTTCTGAATAGCTGTGGGGTC	10	0.3	2.625	50
ABHD9	CCCCCATGTGCGGTGTACCAA	11	CATAGACAGCAGCTTCTCGGGC	12	0.3	2.625	20
ARHGEF6	GTACAGCTGCTCTGGAAGAGGATG	13	TTCCTCAGGGAGTAGGACTTGTGG	14	0.3	2.625	50
ATP2A2	GCCTGCCACCAATCCTGC	15	CCAGATCACCCAGGGGCATTATG	16	0.3	3	50
ATP6V1C2	CGCATAAGAAAGTCAATCCACCAAGCG	17	CTTGGTTATTGAGTTGCAGTCCCGG	18	0.3	3	10
CLEC3B	TGAGGAGCTCAAGAGCCGT	19	GGAAGGTCTTCGTCGTGGTGA	20	0.3	3	50
CYP3A5	AGTTTGGCCCAATAAGCCACCAC	21	GGAATCACCCACCATGACCCCTTGG	22	0.3	3	20
DNAJC22	AGGAGTGGCGAAGCTCTATGAG	23	ACTCCGATGTTATTTCTCATTTGTGCCC	24	0.3	3	5
GPM6B	TATCACCTGTTTCATTTGGCCTGTG	25	TTTAGTGCAGCAATCTTCCGACTC	26	0.3	3	50
HIST3H2A	CCGTACTGTCTGCCCAAGAGAC	27	GAGCCTTATGTCCATCAAGGGGGC	28	0.3	2.625	10
IGSF9	GACAGTCAGCAAGAGGAGGAACAC	29	CCAAGCCCAGGTGCAGAGTC	30	0.3	2.625	5
KANK4	AGCAGCCTGACTGACAAGGC	31	GCTCTCAGAAGCCCGCAATTTC	32	0.3	2.625	5
KIT	GGAAGGCTTCCGGATGCTCAG	33	GATTGGTCTCTCTGAAATCTGCTTCTC	34	0.3	3	10
LRP11	CACGTCTGCAGGGGGA	35	CACGTTGTCAAGAGCTTCTCTGC	36	0.3	2.625	50
MELK	CAGAGACGGGCCAGAAAGAC	37	TGCATTGTCACTTCCCAAATCTGACTG	38	0.3	3	10
METTL7B	GGCTGATGGCTCCATGGATG	39	TCCATATGGTTCTGCCACATGCT	40	0.3	2.625	20
MF12 isoform 2	GAAGATTGGTAAGAGCCCAAGGG	41	ATCAGCCACTCCCCGCTTCTC	42	0.3	2.625	1
MF12 isoforms 1 & 2	GTGCTGTGCTTCACAAAAGCCAC	43	GACAAGAGCCCCCTGGAGAG	44	0.3	2.625	5
MF12 isoform 1	CTCACTGCTGTGAGGACGTCAC	45	CCTTCACCTTGGATGAGCTTCGG	46	0.3	2.625	5
P2RY14	ACTTGGGCCACTTCAAGACGAC	47	CATCTTGTAACTTCTGAAGGCAGAGGC	48	0.3	3	5
PIK3R1	CACCTTCTGTCCGGGAGAGCAG	49	GGCAAAGCCATAGCCAGTTGC	50	0.3	3	50
PKIA	TCCTCTGCAAGTGGCAACAGC	51	CTGGGCTTCCCACTTTGT	52	0.3	2.625	10
REXO1L1	ACCGGGAGTTCAGTTGATTCC	53	GCAGACTGGGTGGGACTTCTC	54	0.3	2.625	1
RXFP1	AGTGCCATAACAGCAACTGAAATACGG	55	CACATCCCAAGAGGTTATGGTACCTGG	56	0.3	3	50
UBE2C	CCTTGAACACACATGCTGCCGA	57	ACAAGGACAGGCTGGGCAG	58	0.15	3	1

Table 8. Log-rank test p-values and ranking for all risk categories in the discovery set 2

Model	Risk Category		Ranking	
	Median	Tertile	Median based	Tertile based
ABHD9	0.486944	0.052072	17	15
CLEC3B	0.178308	0.101169	15	16
HIST3H2A	0.298723	0.744647	16	21
LRP11	0.714154	0.316728	19	20
METTL7B	0.137747	0.015094	14	10
ARHGEF6	0.036545	0.024611	10	11
ATP2A2	0.800665	0.229129	21	19
GPM6B	0.003968	0.038667	3	13
PIK3R1	0.047408	0.035071	11	12
RXFP1	0.013167	0.005214	7	4
CYP3A5	0.002917	0.013174	1	8
KIT	0.068266	0.014172	12	9
PKIA	0.119672	0.039476	13	14
KANK4	0.025092	0.004116	9	3
P2RY14	0.790573	0.103643	20	17
IGSF9	0.519312	0.799465	18	22
ATP6V1C2	0.920056	0.174493	22	18
MELK	0.003316	0.008656	2	6
All genes	0.008861	9.4411e-06	5	1
Backward	0.010535	0.011296	6	7
Forward	0.022636	0.002877	8	2
Penalized	0.004233	0.006174	4	5

Validation set

[0059] 206 samples and 18 genes were available for the biological validation. Mean follow-up time for the cohort was almost 4 years (1407 days). Patient characteristics were similar to the discovery cohort (**Table 1**). Continuous risk scores were calculated with
5 qPCR-based gene expression values of the validation set using coefficients and cutoffs of the six best risk models obtained in the discovery set. Kaplan-Meier analyses of overall survival according to the six risk categories are presented in **Figure 2**.

[0060] Three risk models were validated (p-value < 0.05, **Figure 2**). The single gene model *GPM6B* and the two gene model *METTL7B+KANK4* were associated with survival.
10 However, the most strongly associated with overall survival was the single gene model based on the expression level of *KANK4* (p-value < 0.005). In this model, survival curves were not different between the average risk and low-risk individuals (**Figure 2**), suggesting that the qPCR test can delineate high risk individuals from the others, but not low-risk and average-risk individuals. iAUROC of *KANK4*-based risk categories (0.63) and
15 pathologic staging (0.54) were compared to test if the new risk score was better than pathologic staging to predict overall survival. Wilcoxon rank sum test p-value was smaller than 1E-5, indicating that *KANK4* gene expression values are better at predicting survival than pathologic staging.

[0061] Multivariate and univariate analyses of covariates previously associated with
20 overall survival of lung cancer is presented in **Table 6**. When covariates were taken separately or together, risk categories derived from *KANK4* expression values and age were associated with survival (p-value < 0.05). By univariate analysis, the low-risk and average-risk categories were not significantly different (p-value of 0.31).

Example 4

25 Discussion

[0062] In this study, we identified a prognostic gene expression signature in non-tumor lung specimens that can differentiate patients with early-stage NSCLC beyond clinicopathology staging in order to guide chemotherapy. More specifically, the mRNA expression levels of 22 genes measured in non-tumor lung tissues were found as
30 potential biomarkers to predict overall survival following lung resection for stage I adenocarcinoma. From this discovery, we aim at developing a qPCR assay with optimal

discriminatory performance. We identified many models strongly associated with survival that included the expression levels of all genes, selected combinations of genes, and single gene models. The top six models found in the discovery set were locked down and evaluated in an independent set of samples. Three models were validated and the most
5 strongly replicated model was based of the mRNA expression level of a single gene, namely KN motif and ankyrin domain 4 (KANK4). The KANK4-qPCR assay is urgently needed in clinic to improve risk-stratification in patients undergoing surgery for stage I adenocarcinoma.

[0063] The biological function of *KANK4* is largely unknown. The gene was first
10 identified with other members of the family based on domain and phylogenetic analyses with *KANK1*³⁶. *KANK1* was originally identified as a tumor suppressor gene in renal cell carcinoma³⁷. The Kank family of proteins consists of four members, Kank1-Kank4³⁸. The biological function of *KANK4* remains to be elucidated, but the gene was found strongly expressed in normal human lung tissues³⁶. The low level of expression associated with
15 poor prognostic following lung tumor resection observed in the current study suggests that *KANK4* may act as a tumor suppressor gene akin to *KANK1* in renal cell carcinoma.

[0064] Previous studies have identified gene expression signature in lung tumor that predict survival following lung cancer surgery⁴⁻²⁴. Our expression signature was developed on a mature genomic technology, i.e. testing all known genes. Selection of patients were
20 clearly defined in terms of subtype and stage (i.e. stage I adenocarcinoma). More importantly, we studied non-tumor lung tissues, instead of tumors. Tumors are heterogeneous biological materials consisting of multiple clones that are genetically distinct²⁵. Obtaining a robust gene expression signature from tumors is likely to be very challenging. Non-tumor lung tissues harvested from a site distant from the tumor during
25 the pathological examination are more representative of tissues that remain in the patients after surgery. We thus believe that non-tumor lung specimens are more likely to have a molecular signature that predicts relapse.

[0065] In Canada, more than 25,000 new cases of lung cancer are diagnosed yearly. For patients, surgical resection is the only hope of cure but recurrence rate following
30 resection is very high (35-50% in 5 years)³. This recurrence rate is high compared to resection of other common solid tumors such as colon and breast cancer. Little progress has been made during the past 30 years to reduce disease recurrence and subsequent

mortality after lung tumor resection. Adjuvant platinum-based chemotherapy is now a proven therapy to improve survival in advance-stages lung tumors. However, no data support the use of this chemotherapy in patients with early-stages tumors. Our new qPCR assay based on the expression of a single gene allows delineation of high-risk group. Only
5 in Canada, this assay has the potential to expedite the clinical decision-making process for approximately 5,000 patients. Adjuvant chemotherapy in the high risk group of patients undergoing lung cancer resection for stage I adenocarcinoma has the potential to prevent relapse and save social and economic cost associated with second diagnosis and treatment.

10 **[0066]** In conclusion, we are developing a qPCR prognostic gene expression signature to guide adjuvant chemotherapy following surgical resection of lung tumor for patients with stage I adenocarcinoma. This prognostic test responds to an urgent medical need in order to reduce the socio-economic burden of lung cancer. Large-scale multi-site interventional clinical trials are needed to fully demonstrate the clinical benefit of the qPCR
15 assay, initiate negotiation with regulatory bodies and payers who properly present credentials and accept reimbursements, and finally ensure entry into the health care system.

[0067] While the invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications and
20 this application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure as come within known or customary practice within the art to which the invention pertains and as may be applied to the essential features hereinbefore set forth, and as follows in the scope of the appended claims.

25 **[0068]** All patents, patent applications and publications mentioned in this specification are herein incorporated by reference to the same extent as if each independent patent, patent application, or publication was specifically and individually indicated to be incorporated by reference.

References

30 1. Herbst, R.S., Heymach, J.V. & Lippman, S.M. Lung cancer. *N Engl J Med* **359**, 1367-80 (2008).

2. Edge, S.B. *et al.* *AJCC Cancer Staging Manual*, 646 (Springer, 2009).
3. Jemal, A., Siegel, R., Xu, J. & Ward, E. Cancer statistics, 2010. *CA Cancer J Clin* **60**, 277-300 (2010).
- 5 4. Subramanian, J. & Simon, R. Gene expression-based prognostic signatures in lung cancer: ready for clinical use? *J Natl Cancer Inst* **102**, 464-74 (2010).
5. Kratz, J.R. *et al.* A practical molecular assay to predict survival in resected non-squamous, non-small-cell lung cancer: development and international validation studies. *Lancet* **379**, 823-32 (2012).
- 10 6. Boutros, P.C. *et al.* Prognostic gene signatures for non-small-cell lung cancer. *Proc Natl Acad Sci U S A* **106**, 2824-8 (2009).
7. Chen, H.Y. *et al.* A five-gene signature and clinical outcome in non-small-cell lung cancer. *N Engl J Med* **356**, 11-20 (2007).
8. Lau, S.K. *et al.* Three-gene prognostic classifier for early-stage non small-cell lung cancer. *J Clin Oncol* **25**, 5562-9 (2007).
- 15 9. Mostert, W. *et al.* Age- and sex-specific genomic profiles in non-small cell lung cancer. *JAMA* **303**, 535-43 (2010).
10. Navab, R. *et al.* Prognostic gene-expression signature of carcinoma-associated fibroblasts in non-small cell lung cancer. *Proc Natl Acad Sci U S A* **108**, 7160-5 (2011).
- 20 11. Shedden, K. *et al.* Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med* **14**, 822-7 (2008).
12. Zhu, C.Q. *et al.* Prognostic and predictive gene signature for adjuvant chemotherapy in resected non-small-cell lung cancer. *J Clin Oncol* **28**, 4417-24 (2010).
13. Roepman, P. *et al.* An immune response enriched 72-gene prognostic profile for early-stage non-small-cell lung cancer. *Clin Cancer Res* **15**, 284-90 (2009).
- 25 14. Sun, Z., Wigle, D.A. & Yang, P. Non-overlapping and non-cell-type-specific gene expression signatures predict lung cancer survival. *J Clin Oncol* **26**, 877-83 (2008).
15. Skrzypski, M. *et al.* Three-gene expression signature predicts survival in early-stage squamous cell carcinoma of the lung. *Clin Cancer Res* **14**, 4794-9 (2008).
- 30 16. Raz, D.J. *et al.* A multigene assay is prognostic of survival in patients with early-stage lung adenocarcinoma. *Clin Cancer Res* **14**, 5565-70 (2008).
17. Larsen, J.E. *et al.* Expression profiling defines a recurrence signature in lung squamous cell carcinoma. *Carcinogenesis* **28**, 760-6 (2007).
18. Larsen, J.E. *et al.* Gene expression signature predicts recurrence in lung adenocarcinoma. *Clin Cancer Res* **13**, 2946-54 (2007).
- 35 19. Guo, L. *et al.* Constructing molecular classifiers for the accurate prognosis of lung adenocarcinoma. *Clin Cancer Res* **12**, 3344-54 (2006).
20. Potti, A. *et al.* A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer. *N Engl J Med* **355**, 570-80 (2006).

21. Lu, Y. *et al.* A gene expression signature predicts survival of patients with stage I non-small cell lung cancer. *PLoS Med* **3**, e467 (2006).
22. Raponi, M. *et al.* Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung. *Cancer Res* **66**, 7466-72 (2006).
- 5 23. Tomida, S. *et al.* Gene expression-based, individualized outcome prediction for surgically treated lung cancer patients. *Oncogene* **23**, 5360-70 (2004).
24. Beer, D.G. *et al.* Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med* **8**, 816-24 (2002).
- 10 25. Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* **366**, 883-92 (2012).
26. Spira, A. *et al.* Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat Med* **13**, 361-6 (2007).
27. Aberle, D.R. *et al.* Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* **365**, 395-409 (2011).
- 15 28. Edge, S.B. *et al.* *AJCC Cancer Staging Manual*, 646 (Springer, New York, NY, 2010).
29. Irizarry, R.A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249-64 (2003).
30. Heber, S. & Sick, B. Quality assessment of Affymetrix GeneChip data. *OMICS* **10**, 358-68 (2006).
- 20 31. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Royal Stat Soc Ser B* **57**, 289-300 (1995).
32. R Core Team. R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria* (2013).
33. Therneau, T. A Package for Survival Analysis in S+. *R package version 2.37-4* (2013).
- 25 34. Haibe-Kains, B., Desmedt, C., Sotiriou, C. & Bontempi, G. A comparative study of survival models for breast cancer prognostication based on microarray data: does a single gene beat them all? *Bioinformatics* **24**, 2200-8 (2008).
35. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw* **33**, 1-22 (2010).
- 30 36. Zhu, Y., Kakinuma, N., Wang, Y. & Kiyama, R. Kank proteins: a new family of ankyrin-repeat domain-containing proteins. *Biochim Biophys Acta* **1780**, 128-33 (2008).
37. Sarkar, S. *et al.* A novel ankyrin repeat-containing gene (Kank) located at 9p24 is a growth suppressor of renal cell carcinoma. *J Biol Chem* **277**, 36585-91 (2002).
- 35 38. Kakinuma, N., Zhu, Y., Wang, Y., Roy, B.C. & Kiyama, R. Kank proteins: structure, functions and diseases. *Cell Mol Life Sci* **66**, 2651-9 (2009).

CLAIMS

1. A method for prognosing a cancer recurrence in a human subject having undergone surgical resection of a non-small cell lung cancer (NSCLC); the method comprising:

5 a) determining a gene expression profile present in a non-tumor lung sample derived from the subject, the gene expression profile comprising measuring the level of expression of at least one gene selected from the group consisting of: **ABHD9** (also 3BHD9 or EPHX3): epoxide hydrolase 3; **CLEC3B**: C-type lectin domain family 3, member B; **HIST3H2A**: histone cluster 3, H2a; **LRP11**: low
10 density lipoprotein receptor-related protein 11; **METTL7B**: methyltransferase like 7B; **ARHGEF6**: Rac/Cdc42 guanine nucleotide exchange factor (GEF) 6; **ATP2A2**: ATPase, Ca⁺⁺ transporting, cardiac muscle, slow twitch 2; **GPM6B**: glycoprotein M6B; **PIK3R1**: phosphoinositide-3-kinase, regulatory subunit 1 (alpha); **RXFP1**: relaxin/insulin-like family peptide receptor 1; **CYP3A5**: cytochrome P450, family 3, subfamily A, polypeptide 5; **KIT**: v-kit Hardy-Zuckerman 4 feline sarcoma viral
15 oncogene homolog; **PKIA**: protein kinase (cAMP-dependent, catalytic) inhibitor alpha; **KANK4**: KN motif and ankyrin repeat domains 4 (entrez # 163782); **P2RY14**: purinergic receptor P2Y, G-protein coupled, 14; **IGSF9**: immunoglobulin superfamily, member 9; **ATP6V1C2**: ATPase, H⁺ transporting, lysosomal 42kDa, V1 subunit C2; and **MELK**: maternal embryonic leucine zipper kinase; and
20

b) comparing said level to a control level of same said gene present in a control sample;

wherein a decreased or increased level of said gene expression relative to the level in the control sample indicates the risk of recurrence in the non-small cell
25 lung cancer in said subject.

2. The method according to claim 1, wherein said gene expression profile is that of **KANK4**.

3. The method according to claim 1, wherein said gene expression profile is that of **GPM6B**.

4. The method according to claim 1, wherein said gene expression profile is that of **CYP3A5**.
5. The method according to claim 1, wherein said gene expression profile is that of **MELK**.
- 5 6. The method according to claim 1, wherein said gene expression profile is that of a combination of **METTL7B** and **KANK4**.
7. The method according to claim 1, wherein said gene expression profile is that of a combination of all 18 genes listed in claim 1.
8. The method according to claim 1, wherein said gene expression profile is
10 determined using a mRNA sample.
9. The method of claim 1, wherein the gene expression profile is normalized with the use of average expression level of one or more reference genes in the sample, said reference gene being selected from the group consisting of: TBP, GAPDH, HRPT1, GUSB and ACTB.
- 15 10. The method of claim 1, wherein the sample is a non-tumor lung tissue.
11. The method of claim 1 further comprising classifying said subject, whereby a difference in the expression of the gene expression profile in said sample is used for classifying the subject with NSCLC into a poor survival group (high-risk), an average risk group or a good survival group (low risk).
- 20 12. The method of claim 11, further comprising classifying said subject, whereby a difference in the expression of the gene expression profile in said sample is used for classifying the subject with NSCLC into a poor survival group (high-risk) or a good survival group (low risk).
13. The method of claim 1, wherein the gene expression profile is determined by a
25 technique selected from the group consisting of: quantitative real-time PCR or microarray analysis.
14. The method of claim 1, wherein the NSCLC is a stage I adenocarcinoma.

15. The method of claim 1, wherein the subject is a human.
16. The method of claim 13, wherein the gene expression profile is determined by quantitative real-time PCR using at least one nucleic acid primer selected from the group consisting of : SEQ ID NO. 11 to SEQ ID NO. 58.
- 5 17. A composition comprising a plurality of isolated nucleic acid sequences, wherein each isolated nucleic acid sequence hybridizes to:
- a) at least one RNA product of one of the genes as defined in any one of claims 1 to 7; and/or
 - b) a nucleic acid complementary to a),
- 10 wherein the composition is used to measure the level of RNA expression of at least one of said genes.
18. A kit for prognosing or classifying a subject with non-small cell lung carcinoma (NSCLC) comprising a set of detection agents capable of determining the expression profile of at least 1 gene in a test sample, wherein the gene is selected from the group
- 15 consisting of: **ABHD9** (also 3BHD9 or EPHX3): epoxide hydrolase 3; **CLEC3B**: C-type lectin domain family 3, member B; **HIST3H2A**: histone cluster 3, H2a; **LRP11**: low density lipoprotein receptor-related protein 11; **METTL7B**: methyltransferase like 7B; **ARHGEF6**: Rac/Cdc42 guanine nucleotide exchange factor (GEF) 6; **ATP2A2**: ATPase, Ca⁺⁺ transporting, cardiac muscle, slow twitch 2; **GPM6B**: glycoprotein M6B; **PIK3R1**:
- 20 phosphoinositide-3-kinase, regulatory subunit 1 (alpha); **RXFP1**: relaxin/insulin-like family peptide receptor 1; **CYP3A5**: cytochrome P450, family 3, subfamily A, polypeptide 5; **KIT**: v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog; **PKIA**: protein kinase (cAMP-dependent, catalytic) inhibitor alpha; **KANK4**: KN motif and ankyrin repeat domains 4 (entrez # 163782); **P2RY14**: purinergic receptor P2Y, G-protein coupled, 14;
- 25 **IGSF9**: immunoglobulin superfamily, member 9; **ATP6V1C2**: ATPase, H⁺ transporting, lysosomal 42kDa, V1 subunit C2; and **MELK**: maternal embryonic leucine zipper kinase.
19. The kit according to claim 18, wherein said detection agent is capable of detecting the expression profile of **KANK4**.

20. The kit according to claim 18, wherein said detection agent is capable of detecting the expression profile of **GPM6B**.
21. The method according to claim 18, wherein said detection agent is capable of detecting the expression profile of **CYP3A5**.
- 5 22. The method according to claim 18, said detection agent is capable of detecting the expression profile of **MELK**.
23. The method according to claim 18, wherein said detection agent is capable of detecting the expression profile of a combination of **METTL7B** and **KANK4**.
24. The method according to claim 18, said detection agent is capable of detecting the
10 expression profile of a combination of all 18 genes listed in claim 1.
25. The kit of claim 23, wherein the detection agent is a nucleic acid primer-probe set wherein the primer is selected from the group consisting of : SEQ ID NO. 11 to SEQ ID NO. 58.
26. A method for treating a human subject being prognosed for cancer recurrence of a
15 non-small cell lung cancer, the method comprising:
- a) determining a gene expression profile present in a non-tumor lung sample derived from the subject, the gene expression profile comprising the level of expression of at least one gene selected from the group consisting of: **ABHD9** (also 3BHD9 or EPHX3): epoxide hydrolase 3; **CLEC3B**: C-type lectin domain family 3, member B; **HIST3H2A**: histone cluster 3, H2a; **LRP11**: low density lipoprotein receptor-related protein 11; **METTL7B**: methyltransferase like 7B;
20 **ARHGEF6**: Rac/Cdc42 guanine nucleotide exchange factor (GEF) 6; **ATP2A2**: ATPase, Ca⁺⁺ transporting, cardiac muscle, slow twitch 2; **GPM6B**: glycoprotein M6B; **PIK3R1**: phosphoinositide-3-kinase, regulatory subunit 1 (alpha); **RXFP1**:
25 relaxin/insulin-like family peptide receptor 1; **CYP3A5**: cytochrome P450, family 3, subfamily A, polypeptide 5; **KIT**: v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog; **PKIA**: protein kinase (cAMP-dependent, catalytic) inhibitor alpha; **KANK4**: KN motif and ankyrin repeat domains 4 (entrez # 163782); **P2RY14**: purinergic receptor P2Y, G-protein coupled, 14; **IGSF9**: immunoglobulin

superfamily, member 9; **ATP6V1C2**: ATPase, H⁺ transporting, lysosomal 42kDa, V1 subunit C2; and **MELK**: maternal embryonic leucine zipper kinase;

and

b) comparing said amount to a control amount of same said gene present in a control sample;

wherein a decreased or increased amount of said gene expression relative to the amount in the control sample indicates that the non-small cell lung cancer will recur in said subject, whereby the prognosed cancer recurrence in the subject will be treated.

10 **27.** The method according to claim 26, wherein said gene expression profile is from **KANK4**.

28. The method according to claim 26, wherein said gene expression profile is from **GPM6B**.

15 **29.** The method according to claim 26, wherein said gene expression profile is from **CYP3A5**.

30. The method according to claim 26, wherein said gene expression profile is from **MELK**.

31. The method according to claim 26, wherein said gene expression profile is a combination of **METTL7B** and **KANK4**.

20 **32.** The method according to claim 26, wherein said gene expression profile is from a combination of all 18 genes listed in claim 26.

33. The method according to claim 26, wherein said gene expression profile is determined using a mRNA sample.

25 **34.** The method of claim 26, wherein the gene expression profile is normalized with the use of average expression level of one or more reference genes in the sample, said

reference gene being selected from the group consisting of: TBP, GAPDH, HRPT1, GUSB and ACTB.

35. The method of claim 26, wherein the sample is a non-tumor lung tissue.

36. The method of claim 1 further comprising classifying said subject, whereby a
5 difference in the expression of the gene profile in said sample is used for classifying the subject with NSCLC into a poor survival group (high-risk), an average survival group (average risk) or a good survival group (low risk) prior to adapting the course of treatment.

37. The method of claim 36 further comprising classifying said subject, whereby a
10 difference in the expression of the gene profile in said sample is used for classifying the subject with NSCLC into a poor survival group (high-risk) or a good survival group (low risk) prior to adapting the course of treatment.

38. The method of claim 36, wherein the gene expression profile is determined by a technique selected from the group consisting of: quantitative real-time PCR or microarray analysis.

15 **39.** The method of claim 36, wherein the NSCLC is a stage I adenocarcinoma.

40. The method of claim 36, wherein the subject is a human.

41. The method of claim 36, wherein the gene expression profile is determined by quantitative real-time PCR using at least one nucleic acid primer selected from the group consisting of : SEQ ID NO. 11 to SEQ ID NO. 58.

Figures

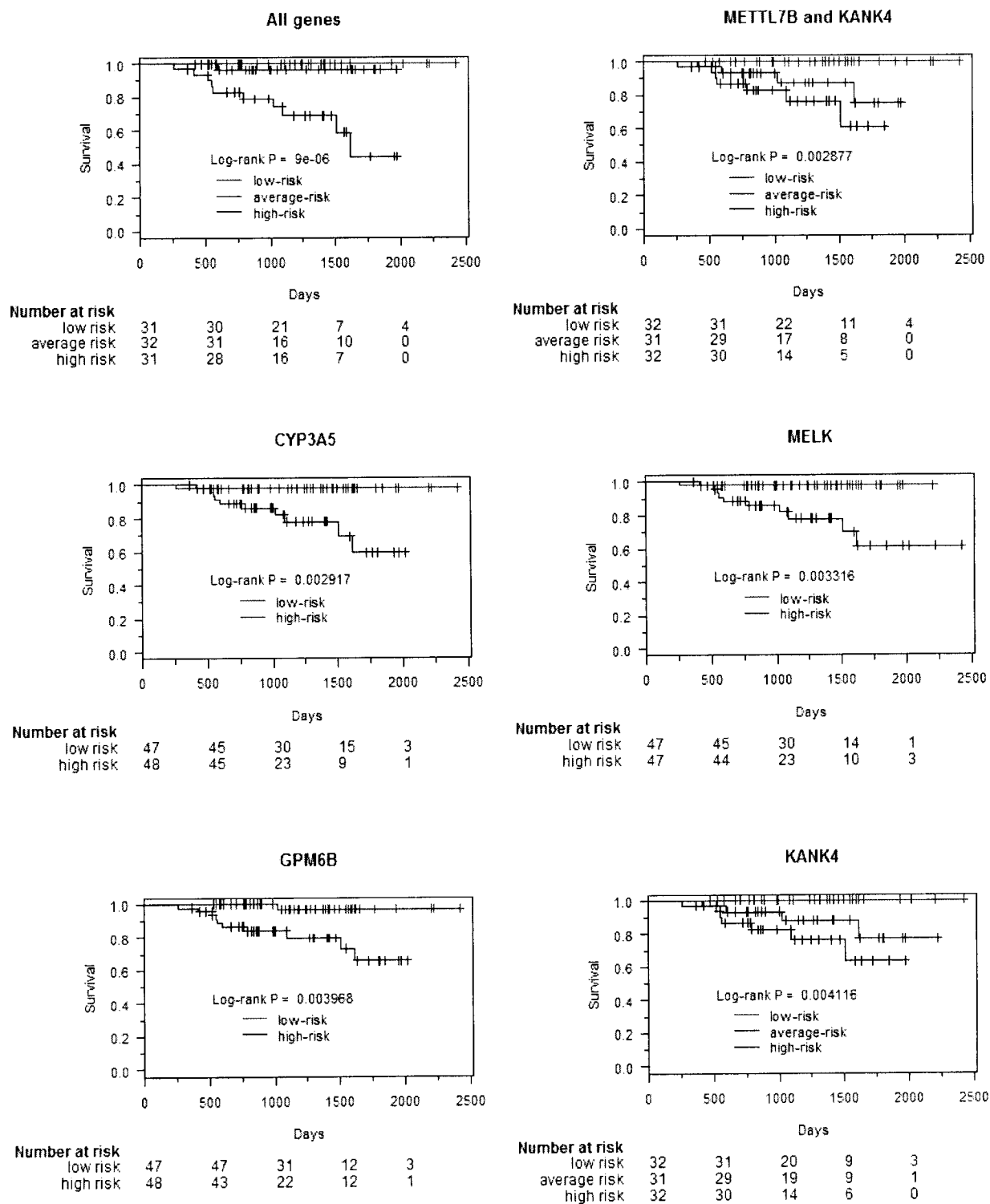


Figure 1

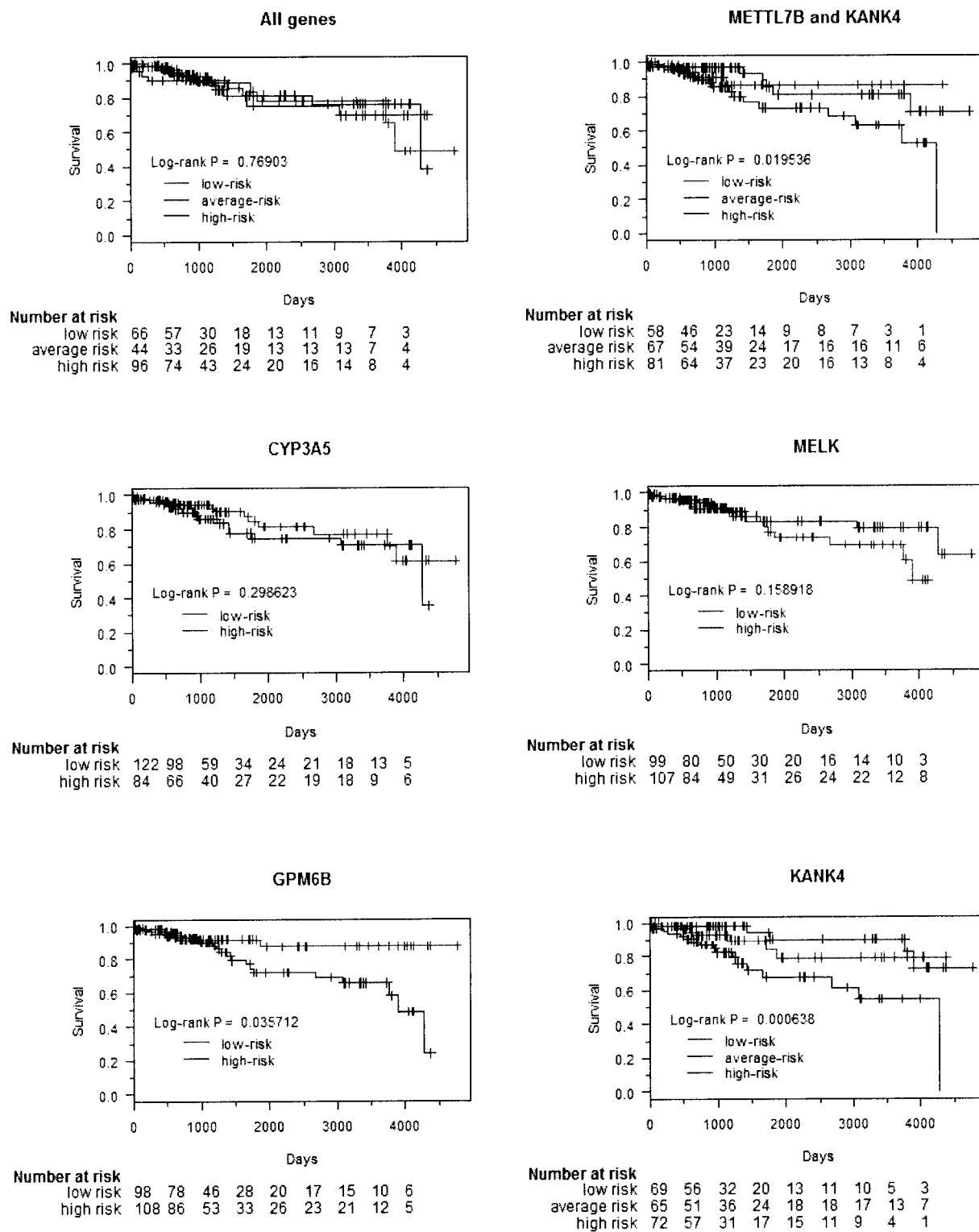


Figure 2

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CA2013/000822

A. CLASSIFICATION OF SUBJECT MATTER IPC: <i>C12Q 1/68</i> (2006.01) , <i>C40B 30/04</i> (2006.01) , <i>C40B 40/06</i> (2006.01) , <i>G01N 33/48</i> (2006.01) According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC: <i>C12Q 1/68</i> (2006.01) , <i>C40B 30/04</i> (2006.01) , <i>C40B 40/06</i> (2006.01) , <i>G01N 33/48</i> (2006.01)		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic database(s) consulted during the international search (name of database(s) and, where practicable, search terms used) Canadian Patent Database, TotalPatent, Medline (key words: cancer, recur, lung, NSCLC, ABHD9, EH3, EPHX3, 3BHD9, gene, expression, signature, predict)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	LU, Y. ET AL. Gene-expression signature predicts postoperative recurrence in stage I non-small cell lung cancer patients. PLoS One. Jan, 2012 (01-2012). Vol. 7, no. 1, e30880. ISSN: 1932-6203, the whole document.	1, 8-18, 25, 26 and 33-41
A	LU, Y. ET AL. A gene expression signature predicts survival of patients with stage I non-small cell lung cancer. PLoS Med. Dec, 2006 (12-2006). Vol. 3, no. 12, e467. ISSN: 1549-1676, the whole document.	1, 8-18, 25, 26 and 33-41
A	MITRA, R. ET AL. Prediction of postoperative recurrence-free survival in non-small cell lung cancer by using an internationally validated gene expression model. Clin Cancer Res. May, 2011 (05-2011). Vol. 17, no. 9, 2934-46. ISSN: 1078-0432 (p) and 1557-3265 (e), the whole document.	1, 8-18, 25, 26 and 33-41
A	PERUMAL, D. ET AL. A novel five gene signature derived from stem-like side population cells predicts overall and recurrence-free survival in NSCLC. PLoS One. Aug, 2012 (08-2012). Vol. 7, no. 8, e43589. ISSN: 1932-6203, the whole document.	1, 8-18, 25, 26 and 33-41
☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.		
*	Special categories of cited documents :	
"A"	document defining the general state of the art which is not considered to be of particular relevance	"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
"E"	earlier application or patent but published on or after the international filing date	"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
"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)	"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
"O"	document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family
"P"	document published prior to the international filing date but later than the priority date claimed	
Date of the actual completion of the international search 16 December 2013 (16-12-2013)		Date of mailing of the international search report 27 December 2013 (27-12-2013)
Name and mailing address of the ISA/CA Canadian Intellectual Property Office Place du Portage I, C114 - 1st Floor, Box PCT 50 Victoria Street Gatineau, Quebec K1A 0C9 Facsimile No.: 001-819-953-2476		Authorized officer Seema Bissoon-Haqqani (819) 934-7932

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CA2013/000822

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	GUO, N. L. ET AL. Confirmation of gene expression-based prediction of survival in non-small cell lung cancer. Clin Cancer Res. Dec, 2008 (12-2008). Vol. 14, no. 24, 8213-20. ISSN: 1078-0432 (p) and 1557-3265 (e), the whole document.	1, 8-18, 25, 26 and 33-41
A	KRATZ, J. R. ET AL. A practical molecular assay to predict survival in resected non-squamous, non-small-cell lung cancer: development and international validation studies. Lancet. Mar, 2012 (03-2012). Vol. 379, no. 9818, 823-32. ISSN: 0140-6736 (p) and 1474-547X (e), the whole document.	1, 8-18, 25, 26 and 33-41
A	SHAO, W. ET AL. The role of gene expression profiling in early-stage non-small cell lung cancer. J Thorac Dis. Jun, 2010 (06-2010). Vol. 2, no. 2, 89-99. ISSN: 2072-1439 (p) and 2077-6624 (e), the whole document.	1, 8-18, 25, 26 and 33-41
A	ZHU, C. Q. ET AL. Understanding prognostic gene expression signatures in lung cancer. Clin Lung Cancer. Sep, 2009 (09-2009). Vol. 10, no. 5, 331-40. ISSN: 1525-7304, the whole document.	1, 8-18, 25, 26 and 33-41
A	ROEPMAN, P. ET AL. An immune response enriched 72-gene prognostic profile for early-stage non-small-cell lung cancer. Clin Cancer Res. Jan, 2009. Vol. 15, no. 1, 284-90. ISSN: 1078-0432 (p) and 1557-3265 (e), the whole document.	1, 8-18, 25, 26 and 33-41
A	HOU, J. ET AL. Gene expression-based classification of non-small cell lung carcinomas and survival prediction. PLoS One. Apr, 2010 (04-2010). Vol. 5, no. 4, e10312. ISSN: 1932-6203, the whole document.	1, 8-18, 25, 26 and 33-41
A	CHEN, D. T. ET AL. Prognostic and predictive value of a malignancy-risk gene signature in early-stage non-small cell lung cancer. J Natl Cancer Inst. Dec, 2011 (12-2011). Vol. 103, no. 24, 1859-70. ISSN: 0027-8874(p) and 1460-2105 (e), the whole document.	1, 8-18, 25, 26 and 33-41
A	LARSEN, J. E. ET AL. Expression profiling defines a recurrence signature in lung squamous cell carcinoma. Carcinogenesis. Mar, 2007 (03-2007). Vol. 28, no. 3, 760-6. ISSN: 1078-0432 (p) and 1557-3265 (e), the whole document.	1, 8-18, 25, 26 and 33-41
A	LARSEN, J. E. ET AL. Gene expression signature predicts recurrence in lung adenocarcinoma. Clin Cancer Res. May, 2007 (05-2007). Vol. 13, no. 10, 2946-54. ISSN: 0143-3334 (p) and 1460-2180 (e), the whole document.	1, 8-18, 25, 26 and 33-41

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of the 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. ☒ Claim Nos. : 26, 33-35 and 37-41

because they relate to subject matter not required to be searched by this Authority, namely :

Although claims 26, 33-35 and 37-41 are directed to a method of treatment of the human/animal body, which this Authority is not obliged to search under Rule 39.1(iv) of the PCT, the search has been carried out based on the alleged effects of the compounds/compositions referred to therein.

2. ☐ Claim 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. ☐ Claim 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 :

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claim 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 claim Nos. :

1, 8-18, 25, 26 and 33-41 (partially as they relate to gene ABHD9)

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.

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

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing filed or furnished:
 - a. (means)
☒ on paper
☒ in electronic form
 - b. (time)
☒ in the international application as filed
☒ together with the international application in electronic form
☐ subsequently to this Authority for the purposes of search
2. ☒ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.
3. Additional comments :