



- (51) **International Patent Classification:**
C12N 15/62 (2006.01)
- (21) **International Application Number:**
PCT/EP2013/052183
- (22) **International Filing Date:**
4 February 2013 (04.02.2013)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
12153907.6 3 February 2012 (03.02.2012) EP
- (71) **Applicant:** MAX-PLANCK-GESELLSCHAFT ZUR FÖRDERUNG DER WISSENSCHAFTEN E.V. [DE/DE]; Hofgartenstr. 8, 80539 München (DE).
- (72) **Inventors:** FERNANDEZ-CUESTA, Lynnette; Brauneberger Str. 8, 50969 Koeln (DE). THOMAS, Roman; Sandstr. 101, 53332 Bornheim (DE). FISCHER, Florian; Von Sandt Platz 2, 50679 Koeln (DE). HEUCKMANN, Johannes; Arnulfstrasse 21, 50937 Koeln (DE).
- (74) **Agent:** VOSSIUS & PARTNER; Jürgen Meier, 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:

- without international search report and to be republished upon receipt of that report (Rule 48.2(g))
- with sequence listing part of description (Rule 5.2(a))

(54) **Title:** NOVEL FUSION GENES IN LUNG CANCER

(57) **Abstract:** The present invention relates to a method for diagnosing lung cancer in a patient, wherein the presence and/or expression of at least one fusion gene in a sample derived from said patient is determined, and wherein the presence or expression of said fusion gene is attributed to the presence of lung cancer in said patient. Furthermore, methods for the selection of susceptibility and/or for the identification of (medical) responders versus non-responders are provided. Also methods for the identification of compounds useful in the medical intervention of cancer, in particular lung cancer, are provided. Accordingly, the present invention also relates to diagnostic means as well as to the medical intervention in proliferative disorders, like cancer and in particular lung cancer.



WO 2013/113942 A2

Novel Fusion genes in Lung Cancer

Technical Field

The present invention relates to a method for diagnosing lung cancer in a patient, wherein the presence and/or expression of at least one fusion gene in a sample derived from said patient is determined, and wherein the presence or expression of said fusion gene is attributed to the presence of lung cancer in said patient. Furthermore, methods for the selection of susceptibility and/or for the identification of (medical) responders versus non-responders are provided. Also methods for the identification of compounds useful in the medical intervention of cancer, in particular lung cancer, are provided. Accordingly, the present invention also relates to diagnostic means as well as to the medical intervention in proliferative disorders, like cancer and in particular lung cancer.

Background of the Invention

Cancer is a disease of the genome, initiated by inactivation of tumor suppressor genes by deletion, mutation or epigenetic mechanisms, and the activation or deregulation of genes as a consequence of point mutation, amplification or cytogenetic abnormalities. Most of the well-characterized structural rearrangements exert their action through either deregulation, resulting in the uncontrolled expression of genes in cell cycle regulation or the like.

In haematological disorders, including malignant lymphomas, rearrangements resulting in fusion genes have been observed. These fusions lead to deregulation. Some of these fusions lead to activation of a tyrosine kinase, which can be successfully targeted using small molecule kinase inhibitors. The best-characterized case is BCR-ABL1 fusion in chronic myelogenous leukemia (CML), which is the target of the tyrosine kinase inhibitor, Gleevec (imatinib). Accordingly, CML patients experience sustained clinical remissions when treated with this drug. Thus, identification of fusion genes represents discoveries of relevant therapeutic targets as well as enhances our understanding of the molecular basis of tumorigenesis. Compared with

hematological disorders, knowledge of the karyotypic features of solid tumors of the lung is very poor.

The dynamics of ongoing efforts to fully annotate the genomes of all major cancer types is reminiscent of that of the Human Genome Project. The analysis of somatic gene copy number alterations and gene mutations associated with cancer (both are here referred to as *lesions*) will thus provide the genetic landscape of human cancer in the near future. The medical implications of these endeavors are exemplified by the success of molecularly targeted cancer therapeutics in genetically defined tumors: the ERBB2/Her2-targeted antibody trastuzumab shrinks tumors in women with *ERBB2*-amplified breast cancer (Slamon et al., 2001); the ABL/KIT/ PDGFR inhibitor imatinib induces responses in patients with chronic myeloid leukemia carrying the *BCR/ABL* translocation (Drucker et al., 2001a; Druker et al., 2001b) as well as in gastrointestinal stromal tumors and melanomas bearing (see Hodi et al NEJM 2008) mutations in *KIT* or *PDGFRA* (Heinrich et al., 2003); finally, *EGFR*-mutant lung tumors are highly sensitive to the EGFR inhibitors gefitinib and erlotinib (Lynch et al., 2004; Paez et al., 2004; Pao et al., 2004). Hence, new identified oncogenes are promising targets to identify anti-cancer drugs.

Non-small cell lung cancer (NSCLC) is one of the two main types of lung carcinoma, *non-small cell* (80.4%) and *small-cell* (16.8%) lung carcinoma, the classification being based on histological criteria. The non-small cell lung carcinomas have a similar prognosis and similar management and comprise three sub-types: squamous cell lung carcinoma, adenocarcinoma and large cell lung carcinoma.

The inventors of the present invention have now unexpectedly found that in cells derived from solid tumors of the lung translocations and fusion genes are expressed. They found that these genetic events have a predictive value and hence can also serve as diagnostic markers for lung cancer. Furthermore, some of the fusion genes are associated to the susceptibility/resistance of tumor cells to certain anti-cancer drugs. Even more, as many of the found fusion genes code for chimeric proteins or fusion proteins with enzymatic function, e.g. kinase function, these fusion genes are well suited as targets for the identification of novel compounds for the treatment of lung cancer.

Thus, the technical problem underlying the present invention is the provision of means and methods for the diagnosis of lung cancer and for evaluation of cells, in particular tumor cells, for their susceptibility or responsiveness to anti-cancer treatment as well as the provision of targets for identification of compounds for treatment of lung cancer.

The technical problem is solved by provision of the embodiments characterized in the claims.

Definitions

A “genomic translocation/rearrangement” is a structural variation resulting of a change in position of a chromosomal segment within a genome. Translocations can happen within the same chromosome (intra-chromosomal) or between two different chromosomes (inter-chromosomal). The rearrangement sometimes causes the fusion of one or more genes (fusion gene). This fusion then results a misallocation and may cause altered expression, and/or total or partial disruption of one or more of the genes comprised in said fusion gene. The rearrangement that fuses two genes resulting in the production of an active protein is called “activating fusion gene”. Hence, an activating fusion gene codes for a “fusion protein” with a new or altered activity. In contrast, a rearrangement may also produce fusion gene that does not code for a functional protein but leads to a loss of function of one or more genes involved. Such fusion gene is also referred to herein as “inactivating fusion gene”.

In addition a fusion gene may also be the result of a trans-splicing event. Hence, this fusion is not the result of a genomic event but a transcriptional mechanism. Whilst genomic rearrangements can be observed by cytogenetic techniques on affected cells, trans-splicing can only be detected by transcriptome sequencing.

The terms “chimeric transcript” and “fusion gene” are used interchangeable herein and include activating fusion genes, inactivating fusion genes and trans-spliced genes, e.g. regulatory fusion genes.

Hence, a “fusion gene” as used herein refers to a hybrid gene (or transcript) formed from two previously separate genes or fragments thereof and thus resulting in gene rearrangement so that

the previously separate genes or fragments thereof are transcribed in, or present on a single transcript. Alternatively, the separate genes may undergo rearrangement independently before they fuse to each other. Accordingly "fused gene" may be construed accordingly to refer to any such rearrangement event. Fused genes can occur as the result of mutations such as
5 translocation (intra-chromosomal rearrangements or inter-chromosomal rearrangements), deletion, inversion, amplification and/or insertion. Moreover fusion genes may also occur as a result of post-transcriptional rearrangements or modifications, like trans-splicing or the like. The resulting fusion gene causes a misallocation, and/or altered expression, and/or total or partial disruption of one or more of the genes comprised in said fusion gene. A fusion gene
10 according to the present invention comprises one or more nucleotide sequences encoding a protein or a fragment of the protein.

By "encoding a protein" it is meant that said nucleotide sequence is derived from a gene that encodes the recited proteins in its wild-type form. However, the nucleotide sequence does not
15 have necessarily be in frame in said fusion gene, i.e. the protein translated from said fusion gene does not have the sequence of the protein encoded by the gene in its wild-type form. This is especially the case if the gene is not the gene at the 5'-end of the fusion gene. In such case due to a fusion which is not in frame with the reading frame of the gene at the 5'-end of the fusion gene, the translation of the nucleotide sequence leads to a protein with another sequence
20 and/or to a premature termination of translation. Furthermore, in context with the present invention the nucleotide sequence encoding a protein or a fragment of the protein also refers to regulatory elements of the transcript of the gene encoding said protein. For example it also includes 5'- and 3'UTR regions. In this case for example the 3'UTR of the gene at the 5'-end of the translocation may be exchanged by the 3'UTR of another gene. In this case the protein
25 encoded by the translocation is the same as the protein encoded by the gene at the 5'-end of the translocation, but its translation is regulated differently.

"Translocation" of genes results in a chromosome abnormality caused by rearrangement of parts between non-homologous chromosomes or between regions within the same
30 chromosome. It is detected on cytogenetics or a karyotype of affected cells. "Deletions" in chromosomes may be of one or more entire genes or only a portion of one or more genes. Genetic "insertion" is the addition of one or more nucleotide base pairs into a genetic sequence.

This can often happen in microsatellite regions due to the DNA polymerase slipping. An "inversion" is rearrangement of genes in a chromosome in which a segment of a gene is reversed end to end. An "amplification" results when a DNA is amplified resulting in the gain in copy number.

5

"Expression" refers to transcription and translation occurring within a host cell. The level of expression of a DNA molecule (a gene) in a cell, tissue, sample or cell culture may be determined on the basis of either the amount of corresponding mRNA that is present within the cell or the amount of the DNA-encoded protein produced by the host cells. Further detail for the term "expression" within the context of the present invention can be obtained via a review of Sambrook *et al.*

10

In accordance with the present invention, the terms "homology" or "percent homology" or "identical" or "identity" or "percent identity" or "percentage identity" or "sequence identity" in the context nucleic acid or protein sequences refers to sequences or subsequences that are the same, or that have a specified percentage of nucleotides that are the same (preferably at least 40% identity, more preferably at least 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97% or 98% identity, most preferably at least 99% identity), when compared and aligned for maximum correspondence over a window of comparison, or over a designated region as measured using a sequence comparison algorithm as known in the art, or by manual alignment and visual inspection. The comparison or alignment may be performed over the full length of a sequence as provided and to be used herein. Hence, if it is referred to a nucleotide sequence or protein sequence having at least 70% identity to a given sequence it will be understood that in preferred embodiments the nucleotide sequence and/or protein sequence has at least 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97% or 98% identity, most preferably at least 99% identity. The nucleic acid provided and to be used in accordance with the present invention relates to a nucleic acid encoding a sequence which is at least 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99 % identical to the protein sequence of SOS1-ADCY3 (SEQ ID NO. 10). The nucleic acid provided and to be used in accordance with the present invention relates to a nucleic acid encoding a sequence which is at least 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99 % identical to the protein sequence of ZNF142-PTK6 (SEQ ID NO. 20).

20

25

30

The nucleic acid provided and to be used in accordance with the present invention relates to a nucleic acid encoding a sequence which is at least 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99 % identical to the protein sequence of FNDC3B-TNIK (SEQ ID NO. 49). The nucleic acid provided and to be used in accordance with the present invention relates to a nucleic acid comprising a sequence which is at least 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99 % identical to the sequence of SOS1-ADCY3 (SEQ ID NO. 9). The nucleic acid provided and to be used in accordance with the present invention relates to a nucleic acid comprising a sequence which is at least 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99 % identical to the sequence of ZNF142-PTK6 (SEQ ID NO. 19). The nucleic acid provided and to be used in accordance with the present invention relates to a nucleic acid comprising a sequence which is at least 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99 % identical to the sequence of FNDC3B-TNIK (SEQ ID NO. 48).

The protein provided and to be used in accordance with the present invention relates to a protein having a sequence with at least at least 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99 % identity to the sequence of SOS1-ADCY3 (SEQ ID NO. 10). The protein provided and to be used in accordance with the present invention relates to a protein having a sequence with at least at least 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99 % identity to the sequence of ZNF142-PTK6 (SEQ ID NO. 20). The protein provided and to be used in accordance with the present invention relates to a protein having a sequence with at least at least 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99 % identity to the sequence of FNDC3B-TNIK (SEQ ID NO. 49). The sequence identities of the nucleic acids and proteins defined above shall also apply to the means, methods and kits sections of the present application.

The sequence may also be completely identical. Sequences having, for example, 70% to 90% or greater sequence identity may be considered to be substantially identical. Such a definition also applies to the complement of a test sequence. Preferably the described identity exists over a region that is at least about 15 to 25 nucleotides in length, more preferably, over a region that is at least about 50 to 100 nucleotides in length and most preferably, over a region that is at least about 800 to 1200 nucleotides in length. Those having skill in the art will know how to determine percent identity between/among sequences using, for example, algorithms such as those based on CLUSTALW computer program (Thompson Nucl. Acids Res. 2 (1994), 4673-

4680) or FASTDB (Brutlag Comp. App. Biosci. 6 (1990), 237-245), as known in the art.

Although the FASTDB algorithm typically does not consider internal non-matching deletions or additions in sequences, i.e., gaps, in its calculation, this can be corrected manually to avoid an overestimation of the % identity. CLUSTALW, however, does take sequence gaps into account in its identity calculations. Also available to those having skill in this art are the BLAST and BLAST 2.0 algorithms (Altschul, (1997) Nucl. Acids Res. 25:3389-3402; Altschul (1993) J. Mol. Evol. 36:290-300; Altschul (1990) J. Mol. Biol. 215:403-410). The BLASTN program for nucleic acid sequences uses as defaults a word length (W) of 11, an expectation (E) of 10, M=5, N=4, and a comparison of both strands. The BLOSUM62 scoring matrix (Henikoff (1989) PNAS 89:10915) uses alignments (B) of 50, expectation (E) of 10, M=5, N=4, and a comparison of both strands.

In order to determine whether a nucleotide residue in a nucleic acid sequence or an amino acid corresponds to a certain position in the nucleotide sequence or protein sequence recited herein, the skilled person can use means and methods well-known in the art, e.g., alignments, either manually or by using computer programs such as those mentioned herein. For example, BLAST 2.0, which stands for Basic Local Alignment Search Tool BLAST (Altschul (1997), loc. cit.; Altschul (1993), loc. cit.; Altschul (1990), loc. cit.), can be used to search for local sequence alignments. BLAST, as discussed above, produces alignments of nucleotide sequences to determine sequence similarity. Because of the local nature of the alignments, BLAST is especially useful in determining exact matches or in identifying similar sequences. The fundamental unit of BLAST algorithm output is the High-scoring Segment Pair (HSP). An HSP consists of two sequence fragments of arbitrary but equal lengths whose alignment is locally maximal and for which the alignment score meets or exceeds a threshold or cut-off score set by the user. The BLAST approach is to look for HSPs between a query sequence and a database sequence, to evaluate the statistical significance of any matches found, and to report only those matches, which satisfy the user-selected threshold of significance. The parameter E establishes the statistically significant threshold for reporting database sequence matches. E is interpreted as the upper bound of the expected frequency of chance occurrence of an HSP (or set of HSPs) within the context of the entire database search. Any database sequence whose match satisfies E is reported in the program output.

Analogous computer techniques using BLAST (Altschul (1997), loc. cit.; Altschul (1993), loc. cit.; Altschul (1990), loc. cit.) are used to search for identical or related molecules in nucleotide databases such as GenBank or EMBL. This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score which is defined as:

$$\frac{\% \text{ sequence identity} \times \% \text{ maximum BLAST score}}{100}$$

and it takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1-2% error; and at 70, the match will be exact. Similar molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules. Another example for a program capable of generating sequence alignments is the CLUSTALW computer program (Thompson (1994) Nucl. Acids Res. 2:4673-4680) or FASTDB (Brutlag (1990) Comp. App. Biosci. 6:237-245), as known in the art.

Oligonucleotides, such as primers, or polynucleotides may be prepared using any suitable method, such as, for example, the phosphotriester and phosphodiester methods or automated embodiments thereof. In one such automated embodiment diethylphosphoramidites are used as starting materials and may be synthesized as described by Beaucage et al., Tetrahedron Letters, 22:1859-1862 (1981), which is hereby incorporated by reference. One method for synthesizing oligonucleotides on a modified solid support is described in U.S. Pat. No. 4,458,006, which is hereby incorporated by reference. It is also possible to use a primer which has been isolated from a biological source (such as a restriction endonuclease digest). Preferred primers have a length of from about 15-100, more preferably about 20-50, most preferably about 20-40 bases.

As used herein, a kit is a packaged combination optionally including instructions for use of the combination and/or other reactions and components for such use.

The meaning of the terms “cell”, “tissue” and “cell culture” is well known in the art and may, for example, be deduced from “The Cell” (Garland Publishing, Inc.). Generally, the term “cell” used herein refers to a single cell or a plurality of cells. The term “plurality of cells” means in the context of the present invention a group of cells comprising more than a single cell. Thereby, the cells out of said group of cells may have a similar function. Said cells may be connected cells and/or separate cells. The term “tissue” in the context of the present invention particularly means a group of cells that perform a similar function. The term “cell culture” means in context of the present invention cells as defined herein above which are grown/cultured under controlled conditions. Cell culture(s) comprise in particular cells (derived/obtained) from multicellular eukaryotes, preferably animals as defined elsewhere herein. It is to be understood that the term “cell culture” as used herein refers also “tissue culture” and/or “organ culture”, an “organ” being a group of tissues which perform the same function.

Preferably, the cell, tissue or cell culture to be selected comprise/are derived from or are a tumor cell, preferably a lung tumor or lung tumor cell. The tumor cells may, for example, be obtained from a biopsy, in particular a biopsy/biopsies from a patient/subject suffering from or being prone to suffering from lung cancer. It is preferred herein that said subject is a human.

The term “mammalian tumor cell” used herein refers to tumor cell which is derived from or is a tumor cell from a mammal, the term mammal being defined herein below. As described herein above in respect of “cell”, “tissue” and “cell culture” the “mammalian tumor cells” or cancer cells may be obtained from a biopsy, in particular a biopsy/biopsies from a patient/subject suffering from lung cancer or a patient/subject being prone to suffer from said disorders. The term “tumor cell” also relates to “cancer cells”

Generally, said tumor cell or cancer cell may be obtained from any biological source/organism, particularly any biological source/organism, suffering from or being prone to suffer from the above-mentioned lung cancer.

Preferably, the (tumor) cell(s) or (cancer) cell to be contacted is (are) obtained/derived from an animal. More preferably, said (tumor)/cancer cell(s) is (are) derived from a mammal. The

meaning of the terms “animal” or “mammal” is well known in the art and can, for example, be deduced from Wehner und Gehring (1995; Thieme Verlag). Non-limiting examples for mammals are even-toed ungulates such as sheep, cattle and pig, odd-toed ungulates such as horses as well as carnivores such as cats and dogs. In the context of this invention, it is particularly envisaged that DNA samples are derived from organisms that are economically or scientifically important. Scientifically or experimentally important organisms include, but are not limited to, mice, rats, rabbits, guinea pigs and pigs.

The tumor cell(s) may also be obtained from primates, which comprise lemurs, monkeys and apes. The meaning of the terms “primate”, “lemur”, “monkey” and “ape” is known and may, for example, be deduced by an artisan from Wehner und Gehring (1995, Thieme Verlag). As mentioned above, the tumor or cancer cell(s) is (are) most preferably derived from a human being suffering from lung cancer, e.g. non-small cell lung cancer. In context of this invention particular useful cells, in particular tumor or cancer cells, are, accordingly, human cells. These cells can be obtained from e.g. biopsies or from biological samples but the term “cell” also relates to *in vitro* cultured cells. In accordance with the present invention, the methods provided herein may comprise a step of obtaining a sample from an organism, such as a human. The definitions and explanations given herein in relation to a “sample” (e.g. a biological sample, like a biopsy) apply, *mutatis mutandis*, in this context.

Protein detection assays (immunoassays) can be of any type applied in the field of diagnostics, including but not restricted to assays methods based on enzymatic reactions, luminescence, fluorescence, radiochemicals. Preferred detection methods comprise strip tests, radioimmunoassay, chemiluminescence- and fluorescence- immunoassay, Immunoblot assay, Enzyme-linked immunoassay (ELISA), Luminex-based bead arrays, and protein microarray assay. The assay types can further be microtitre plate-based, chip-based, bead-based, wherein the biomarker proteins can be attached to the surface or in solution. The assays can be homogenous or heterogeneous assays, sandwich assays, competitive and non-competitive assays (The Immunoassay Handbook, Ed. David Wild, Elsevier LTD, Oxford; 3rd ed. (May 2005), ISBN-13: 978-0080445267; Hultschig C et al., Curr Opin Chem Biol. 2006 Feb;10(1):4-10. PMID: 16376134)

Detailed Description of the Invention

5 The above-mentioned problems and needs are solved by the present invention. The inventors unexpectedly found that lung cancer is associated with the presence and/or expression of fusion genes comprising nucleotide sequences encoding which encode proteins of specific families. Hence, in one embodiment the present invention relates to a method for diagnosing lung cancer in a patient comprising the steps of: - determining the presence and/or expression of at least one
10 fusion gene according to the present invention as outlined herein below in a sample derived from said patient; and - attributing the presence or expression of said fusion gene to the presence of lung cancer in said patient.

The present invention relates to a nucleic acid encoding a sequence at least 70% identical to a
15 protein sequence selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 10), ZNF142-PTK6 (SEQ ID NO. 20), and FNDC3B-TNIK (SEQ ID NO. 49).

The present invention relates to a nucleic acid comprising a sequence at least 70% identical to a sequence selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 9), ZNF142-
20 PTK6 (SEQ ID NO. 19), and FNDC3B-TNIK (SEQ ID NO. 48).

The present invention relates to a nucleic acid selected from the group consisting of

- a) a nucleic acid encoding a protein sequence selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 10), ZNF142-PTK6 (SEQ ID NO. 20), and FNDC3B-TNIK
25 (SEQ ID NO. 49);
- b) a nucleic acid comprising a sequence selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 9), ZNF142-PTK6 (SEQ ID NO. 19), and FNDC3B-TNIK (SEQ ID NO. 48);
- c) a nucleic acid encoding a protein sequence having at least 70% identity to a protein
30 sequence selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 10), ZNF142-PTK6 (SEQ ID NO. 20), and FNDC3B-TNIK (SEQ ID NO. 49);
- d) a nucleic acid comprising a sequence having at least 70% identity to a nucleic acid

sequence selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 9), ZNF142-PTK6 (SEQ ID NO. 19), and FNDC3B-TNIK (SEQ ID NO. 48);

e) a nucleic acid hybridizing to the complementary strand of any one of the nucleic acid sequences of a) to d) under stringent hybridizing conditions; and

5 f) a nucleic acid being degenerate as a result of the genetic code to the nucleotide sequence of the nucleic acids of any one of a) to e)

The present invention relates to a protein having a sequence with at least 70 % identity to a sequence selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 10), ZNF142-PTK6 (SEQ ID NO. 20), and FNDC3B-TNIK (SEQ ID NO. 49).

The present invention relates to a protein selected from the group consisting of

a) a protein comprising a sequence selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 10), ZNF142-PTK6 (SEQ ID NO. 20), and FNDC3B-TNIK (SEQ ID NO. 49);

b) a protein comprising a sequence encoded by a nucleic acid having a sequence selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 9), ZNF142-PTK6 (SEQ ID NO. 19), and FNDC3B-TNIK (SEQ ID NO. 48);

c) a protein having at least 70 % identity to a sequence selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 10), ZNF142-PTK6 (SEQ ID NO. 20), and FNDC3B-TNIK (SEQ ID NO. 49);

d) a protein comprising a sequence encoded by a nucleic acid hybridizing to the complementary strand of the nucleic acid sequence as defined in b) under stringent hybridizing conditions; and

25 e) a protein comprising a sequence encoded by a nucleic acid being degenerate as a result of the genetic code to the nucleotide sequence of the nucleic acid as defined in b) or d). .

The present invention relates to a method for diagnosing lung cancer in a patient comprising the steps of:

- 30 - determining the presence and/or expression of at least one fusion gene in a sample derived from said patient; and
- attributing the presence or expression of said fusion gene to the presence of lung cancer

in said patient;

wherein the fusion gene comprises a nucleic acid selected from the group consisting of

- a) a nucleic acid encoding a protein sequence of SOS1-ADCY3 (SEQ ID NO. 10);
- b) a nucleic acid comprising a sequence of SOS1-ADCY3 (SEQ ID NO. 9);
- 5 c) a nucleic acid encoding a protein sequence having at least 70% identity to a protein sequence of SOS1-ADCY3 (SEQ ID NO. 10);
- d) a nucleic acid comprising a sequence having at least 70% identity to a nucleic acid sequence of SOS1-ADCY3 (SEQ ID NO. 9);
- e) a nucleic acid hybridizing to the complementary strand of any one of the nucleic acid sequences of a) to d) under stringent hybridizing conditions; and
- 10 f) a nucleic acid being degenerate as a result of the genetic code to the nucleotide sequence of the nucleic acids of any one of a) to e).

The present invention relates to a method for diagnosing lung cancer in a patient comprising
15 the steps of:

- determining the presence and/or expression of at least one fusion gene in a sample derived from said patient; and
- attributing the presence or expression of said fusion gene to the presence of lung cancer in said patient;

20 wherein the fusion gene comprises a nucleic acid selected from the group consisting of

- a) a nucleic acid encoding a protein sequence of ZNF142-PTK6 (SEQ ID NO. 20);
- b) a nucleic acid comprising a sequence of ZNF142-PTK6 (SEQ ID NO. 19);
- c) a nucleic acid encoding a protein sequence having at least 70% identity to a protein sequence of ZNF142-PTK6 (SEQ ID NO. 20);
- 25 d) a nucleic acid comprising a sequence having at least 70% identity to a nucleic acid sequence of ZNF142-PTK6 (SEQ ID NO. 19);
- e) a nucleic acid hybridizing to the complementary strand of any one of the nucleic acid sequences of a) to d) under stringent hybridizing conditions; and
- f) a nucleic acid being degenerate as a result of the genetic code to the nucleotide
30 sequence of the nucleic acids of any one of a) to e).

The present invention relates to a method for diagnosing lung cancer in a patient comprising

the steps of:

- determining the presence and/or expression of at least one fusion gene in a sample derived from said patient; and
- attributing the presence or expression of said fusion gene to the presence of lung cancer in said patient;

wherein the fusion gene comprises a nucleic acid selected from the group consisting of

- a) a nucleic acid encoding a protein sequence of FNDC3B-TNIK (SEQ ID NO. 49);
- b) a nucleic acid comprising a sequence of FNDC3B-TNIK (SEQ ID NO. 48);
- c) a nucleic acid encoding a protein sequence having at least 70% identity to a protein sequence of FNDC3B-TNIK (SEQ ID NO. 49);
- d) a nucleic acid comprising a sequence having at least 70% identity to a nucleic acid sequence of FNDC3B-TNIK (SEQ ID NO. 48);
- e) a nucleic acid hybridizing to the complementary strand of any one of the nucleic acid sequences of a) to d) under stringent hybridizing conditions; and
- f) a nucleic acid being degenerate as a result of the genetic code to the nucleotide sequence of the nucleic acids of any one of a) to e).

The invention relates to a method of selecting a cell, tissue or cell culture with susceptibility to a MEK inhibitor, comprising the steps:

- determining the presence and/or expression of a fusion gene in said cell, tissue or cell culture, wherein the fusion gene comprises a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase in said cell, tissue or cell culture; and
- attributing the presence and/or expression of said fusion gene to the susceptibility of the cell, tissue or cell culture to a MEK inhibitor,

wherein the fusion gene comprises a nucleic acid selected from the group consisting of

- a) a nucleic acid encoding a protein sequence of SOS1-ADCY3 (SEQ ID NO. 10);
- b) a nucleic acid comprising a sequence of SOS1-ADCY3 (SEQ ID NO. 9);
- c) a nucleic acid encoding a protein sequence having at least 70% identity to a protein sequence of SOS1-ADCY3 (SEQ ID NO. 10);
- d) a nucleic acid comprising a sequence having at least 70% identity to a nucleic acid sequence of SOS1-ADCY3 (SEQ ID NO. 9);

- e) a nucleic acid hybridizing to the complementary strand of any one of the nucleic acid sequences of a) to d) under stringent hybridizing conditions; and
- f) a nucleic acid being degenerate as a result of the genetic code to the nucleotide sequence of the nucleic acids of any one of a) to e).

5

The present invention relates to a method for identifying a responder or non-responder to a MEK inhibitor, the method comprising the steps:

- providing a sample from a patient suspected to suffer or being prone to suffer from lung cancer; and
- 10 - determining the presence and/or expression of a fusion gene comprising a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase, or a fragment thereof;
wherein the presence and/or expression of said fusion gene is indicative for a responder to a MEK inhibitor,
- 15 wherein the fusion gene comprises a nucleic acid selected from the group consisting of
 - a) a nucleic acid encoding a protein sequence of SOS1-ADCY3 (SEQ ID NO. 10);
 - b) a nucleic acid comprising a sequence of SOS1-ADCY3 (SEQ ID NO. 9);
 - c) a nucleic acid encoding a protein sequence having at least 70% identity to a protein sequence of SOS1-ADCY3 (SEQ ID NO. 10);
 - 20 d) a nucleic acid comprising a sequence having at least 70% identity to a nucleic acid sequence of SOS1-ADCY3 (SEQ ID NO. 9);
 - e) a nucleic acid hybridizing to the complementary strand of any one of the nucleic acid sequences of a) to d) under stringent hybridizing conditions; and
 - f) a nucleic acid being degenerate as a result of the genetic code to the nucleotide
- 25 sequence of the nucleic acids of any one of a) to e).

The present invention relates to a method of determining resistance of a cell, tissue or cell culture to an ALK inhibitor, comprising the steps:

- determining the presence and/or expression level of a fusion gene in said cell, tissue or
- 30 cell culture, wherein the fusion gene comprises a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase in said cell, tissue or cell culture;

- comparing the determined expression level to the expression level in a control cell, tissue or cell culture expressing said fusion gene and not being resistant to an ALK inhibitor; and
- attributing a determined expression level higher than in the control cell, tissue or cell culture to the resistance of the cell, tissue or cell culture to a ALK inhibitor; and/or
- attributing a determined expression equal or lower than in the control cell, tissue or cell culture to the susceptibility of the cell, tissue or cell culture to a ALK inhibitor,

wherein the fusion gene comprises a nucleic acid selected from the group consisting of

- a) a nucleic acid encoding a protein sequence of SOS1-ADCY3 (SEQ ID NO. 10);
- b) a nucleic acid comprising a sequence of SOS1-ADCY3 (SEQ ID NO. 9);
- c) a nucleic acid encoding a protein sequence having at least 70% identity to a protein sequence of SOS1-ADCY3 (SEQ ID NO. 10);
- d) a nucleic acid comprising a sequence having at least 70% identity to a nucleic acid sequence of SOS1-ADCY3 (SEQ ID NO. 9);
- e) a nucleic acid hybridizing to the complementary strand of any one of the nucleic acid sequences of a) to d) under stringent hybridizing conditions; and
- f) a nucleic acid being degenerate as a result of the genetic code to the nucleotide sequence of the nucleic acids of any one of a) to e).

The present invention relates to a method for identifying a responder or non-responder to an ALK inhibitor, the method comprising the steps:

- determining the presence/expression level of a fusion gene in a sample from a patient suspected to suffer or being prone to suffer from lung cancer, said fusion gene comprising a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase, or a fragment thereof; and
- comparing the determined expression level to the expression level in a sample from a control patient, wherein the control patient is a responder to ALK inhibitors and expresses said fusion gene;

wherein a determined expression higher than in said sample of a control patient is indicative for a non-responder to a ALK inhibitor; and

wherein a determined expression equal or lower than in said sample of a control patient is indicative for a responder to a ALK inhibitor; and

wherein the fusion gene comprises a nucleic acid selected from the group consisting of

- a) a nucleic acid encoding a protein sequence of SOS1-ADCY3 (SEQ ID NO. 10);
- b) a nucleic acid comprising a sequence of SOS1-ADCY3 (SEQ ID NO. 9);
- c) a nucleic acid encoding a protein sequence having at least 70% identity to a protein
5 sequence of SOS1-ADCY3 (SEQ ID NO. 10);
- d) a nucleic acid comprising a sequence having at least 70% identity to a nucleic acid
sequence of SOS1-ADCY3 (SEQ ID NO. 9);
- e) a nucleic acid hybridizing to the complementary strand of any one of the nucleic acid
sequences of a) to d) under stringent hybridizing conditions; and
- 10 f) a nucleic acid being degenerate as a result of the genetic code to the nucleotide
sequence of the nucleic acids of any one of a) to e).

The present invention relates to a method of selecting a cell, tissue or cell culture with susceptibility to a HAT inhibitor, comprising the steps:

- 15 - determining the presence and/or expression of a fusion gene in said cell, tissue or cell
culture, wherein the fusion gene comprises a nucleotide sequence encoding a protein or
a fragment of the protein, said protein being an adenylyl cyclase in said cell, tissue or
cell culture; and
- attributing the presence and/or expression of said fusion gene to the susceptibility of the
20 cell, tissue or cell culture to a HAT inhibitor,

wherein the fusion gene comprises a nucleic acid selected from the group consisting of

- a) a nucleic acid encoding a protein sequence of SOS1-ADCY3 (SEQ ID NO. 10);
- b) a nucleic acid comprising a sequence of SOS1-ADCY3 (SEQ ID NO. 9);
- c) a nucleic acid encoding a protein sequence having at least 70% identity to a protein
25 sequence of SOS1-ADCY3 (SEQ ID NO. 10);
- d) a nucleic acid comprising a sequence having at least 70% identity to a nucleic acid
sequence of SOS1-ADCY3 (SEQ ID NO. 9);
- e) a nucleic acid hybridizing to the complementary strand of any one of the nucleic acid
sequences of a) to d) under stringent hybridizing conditions; and
- 30 f) a nucleic acid being degenerate as a result of the genetic code to the nucleotide
sequence of the nucleic acids of any one of a) to e).

The present invention relates to a method for identifying a responder or non-responder to a HAT inhibitor, the method comprising the steps:

- determining the presence and/or expression of a fusion gene in a sample from a patient suspected to suffer or being prone to suffer from lung cancer, said fusion gene comprising a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase, or a fragment thereof;

wherein the presence and/or expression of said fusion gene is indicative for a responder to a HAT inhibitor; and

wherein the fusion gene comprises a nucleic acid selected from the group consisting of

- a) a nucleic acid encoding a protein sequence of SOS1-ADCY3 (SEQ ID NO. 10);
- b) a nucleic acid comprising a sequence of SOS1-ADCY3 (SEQ ID NO. 9);
- c) a nucleic acid encoding a protein sequence having at least 70% identity to a protein sequence of SOS1-ADCY3 (SEQ ID NO. 10);
- d) a nucleic acid comprising a sequence having at least 70% identity to a nucleic acid sequence of SOS1-ADCY3 (SEQ ID NO. 9);
- e) a nucleic acid hybridizing to the complementary strand of any one of the nucleic acid sequences of a) to d) under stringent hybridizing conditions; and
- f) a nucleic acid being degenerate as a result of the genetic code to the nucleotide sequence of the nucleic acids of any one of a) to e).

The present invention relates to a method of selecting a cell, tissue or cell culture with susceptibility to a PKA inhibitor, comprising the steps:

- determining the presence and/or expression of a fusion gene in said cell, tissue or cell culture, wherein the fusion gene comprises a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase in said cell, tissue or cell culture; and
- attributing the presence and/or expression of said fusion gene to the susceptibility of the cell, tissue or cell culture to a PKA inhibitor;

wherein the fusion gene comprises a nucleic acid selected from the group consisting of

- a) a nucleic acid encoding a protein sequence of SOS1-ADCY3 (SEQ ID NO. 10);
- b) a nucleic acid comprising a sequence of SOS1-ADCY3 (SEQ ID NO. 9);
- c) a nucleic acid encoding a protein sequence having at least 70% identity to a protein

sequence of SOS1-ADCY3 (SEQ ID NO. 10);

- d) a nucleic acid comprising a sequence having at least 70% identity to a nucleic acid sequence of SOS1-ADCY3 (SEQ ID NO. 9);
- e) a nucleic acid hybridizing to the complementary strand of any one of the nucleic acid sequences of a) to d) under stringent hybridizing conditions; and
- f) a nucleic acid being degenerate as a result of the genetic code to the nucleotide sequence of the nucleic acids of any one of a) to e).

The present invention relates to a method for identifying a responder or non-responder to a PKA inhibitor, the method comprising the steps:

- determining the presence and/or expression of a fusion gene in a sample from a patient suspected to suffer or being prone to suffer from lung cancer, said fusion gene comprising a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase, or a fragment thereof;

wherein the presence and/or expression of said fusion gene is indicative for a responder to a PKA inhibitor; and

wherein the fusion gene comprises a nucleic acid selected from the group consisting of

- a) a nucleic acid encoding a protein sequence of SOS1-ADCY3 (SEQ ID NO. 10);
- b) a nucleic acid comprising a sequence of SOS1-ADCY3 (SEQ ID NO. 9);
- c) a nucleic acid encoding a protein sequence having at least 70% identity to a protein sequence of SOS1-ADCY3 (SEQ ID NO. 10);
- d) a nucleic acid comprising a sequence having at least 70% identity to a nucleic acid sequence of SOS1-ADCY3 (SEQ ID NO. 9);
- e) a nucleic acid hybridizing to the complementary strand of any one of the nucleic acid sequences of a) to d) under stringent hybridizing conditions; and
- f) a nucleic acid being degenerate as a result of the genetic code to the nucleotide sequence of the nucleic acids of any one of a) to e).

The present invention relates to a method of selecting a cell, tissue or cell culture with susceptibility to a PTK inhibitor, comprising the steps:

- determining the presence and/or expression of a fusion gene in said cell, tissue or cell culture, wherein the fusion gene comprises a nucleotide sequence encoding a protein or

a fragment of the protein, said protein being a protein tyrosine kinase, in said cell, tissue or cell culture; and

- attributing the presence and/or expression of said fusion gene to the susceptibility of the cell, tissue or cell culture to a PTK inhibitor,

5 wherein the fusion gene comprises a nucleic acid selected from the group consisting of

- a) a nucleic acid encoding a protein sequence of ZNF142-PTK6 (SEQ ID NO. 20);
- b) a nucleic acid comprising a sequence of ZNF142-PTK6 (SEQ ID NO. 19);
- c) a nucleic acid encoding a protein sequence having at least 70% identity to a protein sequence of ZNF142-PTK6 (SEQ ID NO. 20);
- 10 d) a nucleic acid comprising a sequence having at least 70% identity to a nucleic acid sequence of ZNF142-PTK6 (SEQ ID NO. 19);
- e) a nucleic acid hybridizing to the complementary strand of any one of the nucleic acid sequences of a) to d) under stringent hybridizing conditions; and
- f) a nucleic acid being degenerate as a result of the genetic code to the nucleotide
- 15 sequence of the nucleic acids of any one of a) to e).

The present invention relates to a method for identifying a responder or non-responder to a PTK inhibitor, the method comprising the steps:

- determining the presence and/or expression of a fusion gene in a sample from a patient
- 20 suspected to suffer or being prone to suffer from lung cancer, said fusion gene comprising a nucleotide sequence encoding a protein or a fragment of the protein, said protein being a protein tyrosine kinase, or a fragment thereof;

wherein the presence and/or expression of said fusion gene is indicative for a responder to a PTK inhibitor,

25 wherein the fusion gene comprises a nucleic acid selected from the group consisting of

- a) a nucleic acid encoding a protein sequence of ZNF142-PTK6 (SEQ ID NO. 20);
- b) a nucleic acid comprising a sequence of ZNF142-PTK6 (SEQ ID NO. 19);
- c) a nucleic acid encoding a protein sequence having at least 70% identity to a protein sequence of ZNF142-PTK6 (SEQ ID NO. 20);
- 30 d) a nucleic acid comprising a sequence having at least 70% identity to a nucleic acid sequence of ZNF142-PTK6 (SEQ ID NO. 19);
- e) a nucleic acid hybridizing to the complementary strand of any one of the nucleic acid

sequences of a) to d) under stringent hybridizing conditions; and

- f) a nucleic acid being degenerate as a result of the genetic code to the nucleotide sequence of the nucleic acids of any one of a) to e).

5 The invention relates to a method of selecting a cell, tissue or cell culture with susceptibility to a c-Jun inhibitor, comprising the steps:

- determining the presence and/or expression of a fusion gene in said cell, tissue or cell culture, wherein the fusion gene comprises a nucleotide sequence encoding a protein or a fragment of the protein, said protein being a germinal center kinase, in said cell, tissue
10 or cell culture; and
- attributing the presence and/or expression of said fusion gene to the susceptibility of the cell, tissue or cell culture to a c-Jun inhibitor,

wherein the fusion gene comprises a nucleic acid selected from the group consisting of

- a) a nucleic acid encoding a protein sequence of FNDC3B-TNIK (SEQ ID NO. 49);
- 15 b) a nucleic acid comprising a sequence of FNDC3B-TNIK (SEQ ID NO. 48);
- c) a nucleic acid encoding a protein sequence having at least 70% identity to a protein sequence of FNDC3B-TNIK (SEQ ID NO. 49);
- d) a nucleic acid comprising a sequence having at least 70% identity to a nucleic acid sequence of FNDC3B-TNIK (SEQ ID NO. 48);
- 20 e) a nucleic acid hybridizing to the complementary strand of any one of the nucleic acid sequences of a) to d) under stringent hybridizing conditions; and
- f) a nucleic acid being degenerate as a result of the genetic code to the nucleotide sequence of the nucleic acids of any one of a) to e).

25 The present invention relates to a method for identifying a responder or non-responder to a c-Jun inhibitor, the method comprising the steps:

- determining the presence and/or expression of a fusion gene in a sample from a patient suspected to suffer or being prone to suffer from lung cancer, said fusion gene comprising a nucleotide sequence encoding a protein or a fragment of the protein, said protein being
30 germinal center kinase, or a fragment thereof;

wherein the presence and/or expression of said fusion gene is indicative for a responder to a c-Jun inhibitor,

wherein the fusion gene comprises a nucleic acid selected from the group consisting of

- a) a nucleic acid encoding a protein sequence of FNDC3B-TNIK (SEQ ID NO. 49);
- b) a nucleic acid comprising a sequence of FNDC3B-TNIK (SEQ ID NO. 48);
- c) a nucleic acid encoding a protein sequence having at least 70% identity to a protein
5 sequence of FNDC3B-TNIK (SEQ ID NO. 49);
- d) a nucleic acid comprising a sequence having at least 70% identity to a nucleic acid
sequence of FNDC3B-TNIK (SEQ ID NO. 48);
- e) a nucleic acid hybridizing to the complementary strand of any one of the nucleic acid
sequences of a) to d) under stringent hybridizing conditions; and
- 10 f) a nucleic acid being degenerate as a result of the genetic code to the nucleotide
sequence of the nucleic acids of any one of a) to e).

The present invention relates to a method for identifying compounds for the treatment of lung
cancer comprising the steps:

- 15 - providing a cell expressing a fusion gene, wherein the fusion gene comprises a nucleic
acid selected from the group consisting of
 - a) a nucleic acid encoding a protein sequence of SOS1-ADCY3 (SEQ ID NO. 10);
 - b) a nucleic acid comprising a sequence of SOS1-ADCY3 (SEQ ID NO. 9);
 - c) a nucleic acid encoding a protein sequence having at least 70% identity to a
20 protein sequence of SOS1-ADCY3 (SEQ ID NO. 10);
 - d) a nucleic acid comprising a sequence having at least 70% identity to a nucleic
acid sequence of SOS1-ADCY3 (SEQ ID NO. 9);
 - e) a nucleic acid hybridizing to the complementary strand of any one of the nucleic
acid sequences of a) to d) under stringent hybridizing conditions; and
 - 25 f) a nucleic acid being degenerate as a result of the genetic code to the nucleotide
sequence of the nucleic acids of any one of a) to e);
- contacting said cell to a candidate compound; and
- evaluating the viability of said cell contacted with said candidate compound.

30 The present invention relates to a method for identifying compounds for the treatment of lung
cancer comprising the steps:

- providing a cell expressing a fusion gene, wherein the fusion gene comprises a nucleic

acid selected from the group consisting of

- a) a nucleic acid encoding a protein sequence of ZNF142-PTK6 (SEQ ID NO. 20);
 - b) a nucleic acid comprising a sequence of ZNF142-PTK6 (SEQ ID NO. 19);
 - c) a nucleic acid encoding a protein sequence having at least 70% identity to a protein sequence of ZNF142-PTK6 (SEQ ID NO. 20);
 - d) a nucleic acid comprising a sequence having at least 70% identity to a nucleic acid sequence of ZNF142-PTK6 (SEQ ID NO. 19);
 - e) a nucleic acid hybridizing to the complementary strand of any one of the nucleic acid sequences of a) to d) under stringent hybridizing conditions; and
 - f) a nucleic acid being degenerate as a result of the genetic code to the nucleotide sequence of the nucleic acids of any one of a) to e);
- contacting said cell to a candidate compound; and
 - evaluating the viability of said cell contacted with said candidate compound,

The present invention relates to a method for identifying compounds for the treatment of lung cancer comprising the steps:

- providing a cell expressing a fusion gene, wherein the fusion gene comprises a nucleic acid selected from the group consisting of
 - a) a nucleic acid encoding a protein sequence of FNDC3B-TNFIK (SEQ ID NO. 49);
 - b) a nucleic acid comprising a sequence of FNDC3B-TNFIK (SEQ ID NO. 48);
 - c) a nucleic acid encoding a protein sequence having at least 70% identity to a protein sequence of FNDC3B-TNFIK (SEQ ID NO. 49);
 - d) a nucleic acid comprising a sequence having at least 70% identity to a nucleic acid sequence of FNDC3B-TNFIK (SEQ ID NO. 48);
 - e) a nucleic acid hybridizing to the complementary strand of any one of the nucleic acid sequences of a) to d) under stringent hybridizing conditions; and
 - f) a nucleic acid being degenerate as a result of the genetic code to the nucleotide sequence of the nucleic acids of any one of a) to e);
- contacting said cell to a candidate compound; and
- evaluating the viability of said cell contacted with said candidate compound,

The present invention relates to the use of a cell line expressing at least a fusion gene which comprises a nucleic acid selected from the group consisting of

- a) a nucleic acid encoding a protein sequence of SOS1-ADCY3 (SEQ ID NO. 10);
- b) a nucleic acid comprising a sequence of SOS1-ADCY3 (SEQ ID NO. 9);
- 5 c) a nucleic acid encoding a protein sequence having at least 70% identity to a protein sequence of SOS1-ADCY3 (SEQ ID NO. 10);
- d) a nucleic acid comprising a sequence having at least 70% identity to a nucleic acid sequence of SOS1-ADCY3 (SEQ ID NO. 9);
- e) a nucleic acid hybridizing to the complementary strand of any one of the nucleic acid sequences of a) to d) under stringent hybridizing conditions; and
- 10 f) a nucleic acid being degenerate as a result of the genetic code to the nucleotide sequence of the nucleic acids of any one of a) to e);

for screening and/or validation of a medicament for the treatment of lung cancer.

15 The present invention relates to the use of a cell line expressing at least a fusion gene which comprises a nucleic acid selected from the group consisting of

- a) a nucleic acid encoding a protein sequence of ZNF142-PTK6 (SEQ ID NO. 20);
- b) a nucleic acid comprising a sequence of ZNF142-PTK6 (SEQ ID NO. 19);
- c) a nucleic acid encoding a protein sequence having at least 70% identity to a protein sequence of ZNF142-PTK6 (SEQ ID NO. 20);
- 20 d) a nucleic acid comprising a sequence having at least 70% identity to a nucleic acid sequence of ZNF142-PTK6 (SEQ ID NO. 19);
- e) a nucleic acid hybridizing to the complementary strand of any one of the nucleic acid sequences of a) to d) under stringent hybridizing conditions; and
- 25 f) a nucleic acid being degenerate as a result of the genetic code to the nucleotide sequence of the nucleic acids of any one of a) to e);

for screening and/or validation of a medicament for the treatment of lung cancer.

The present invention relates to the use of a cell line expressing at least a fusion gene which comprises a nucleic acid selected from the group consisting of

- a) a nucleic acid encoding a protein sequence of FNDC3B-TNIIK (SEQ ID NO. 49);

- b) a nucleic acid comprising a sequence of FNDC3B-TNIK (SEQ ID NO. 48);
 - c) a nucleic acid encoding a protein sequence having at least 70% identity to a protein sequence of FNDC3B-TNIK (SEQ ID NO. 49);
 - d) a nucleic acid comprising a sequence having at least 70% identity to a nucleic acid sequence of FNDC3B-TNIK (SEQ ID NO. 48);
 - e) a nucleic acid hybridizing to the complementary strand of any one of the nucleic acid sequences of a) to d) under stringent hybridizing conditions; and
 - f) a nucleic acid being degenerate as a result of the genetic code to the nucleotide sequence of the nucleic acids of any one of a) to e);
- for screening and/or validation of a medicament for the treatment of lung cancer.

The present invention relates to a method for identifying the presence and expression of a fusion gene comprising a nucleic acid encoding an adenylyl cyclase in a sample, the method comprising the step of:

- (i) determining the phosphorylation status of PKA and/or CREB;
 - (ii) comparing the determined phosphorylation status to the phosphorylation status of PKA and/or CREB in a control; and
 - (iii) attributing a higher phosphorylation status of PKA and/or CREB in determined step (i) than in the control to the presence and expression of a fusion gene,
- wherein said fusion gene comprises a nucleic acid selected from the group consisting of
- a) a nucleic acid encoding a protein sequence of SOS1-ADCY3 (SEQ ID NO. 10);
 - b) a nucleic acid comprising a sequence of SOS1-ADCY3 (SEQ ID NO. 9);
 - c) a nucleic acid encoding a protein sequence having at least 70% identity to a protein sequence of SOS1-ADCY3 (SEQ ID NO. 10);
 - d) a nucleic acid comprising a sequence having at least 70% identity to a nucleic acid sequence of SOS1-ADCY3 (SEQ ID NO. 9);
 - e) a nucleic acid hybridizing to the complementary strand of any one of the nucleic acid sequences of a) to d) under stringent hybridizing conditions; and
 - f) a nucleic acid being degenerate as a result of the genetic code to the nucleotide sequence of the nucleic acids of any one of a) to e);

Fusion genes according to the present invention may be grouped into activating, inactivating and regulatory fusion genes:

An activating fusion gene refers to fusion genes in which the fusion partners are fused in frame so that they are expressed with the protein sequence originally encoded by the fusion partners at least partially. For example one or more of the functional domains, e.g. kinase domains, may be present. However, these domains may be modified, e.g. with respect to their selectivity. In another embodiment activating also means that the expression level of at least one fusion partner is increased as compared to the expression in wild type cells.

An inactivating fusion gene refers to a fusion gene wherein the translation of the fusion gene results in a protein that does not exhibit the activity of at least one of the proteins encoded by the fusion partners. For example the nucleotide sequence of one gene comprised in the fusion gene may be fused out of frame, i.e. the reading frame has shifted with respect to the wild type form of the gene. In such case the same nucleotide sequence will be translated into a different protein and/or the translation will stop prematurely due to an early stop codon as consequence of the frame-shift in the sequence.

A regulatory fusion gene refers to a fusion gene wherein the fusion of the fusion partners alters the expression of at least one of the fusion partners, e.g. transcription and/or translation. This may for instance be the case when the 3' untranslated region (3'UTR) of one fusion partner is exchanged by the 3'UTR of another fusion partner resulting in a differently regulated translation. The same holds true for the exchange of 5'UTR or other regulatory elements of mRNA. Such fusion gene may be a consequence of a mechanism of trans-splicing. Fusion genes which are formed by trans-splicing can not be detected on the genomic level but on mRNA or protein level. The transcription-mediated fusion genes may alter the expression of at least one of the fusion partners, e.g. transcription and/or translation.

In a preferred embodiment the fusion gene according to the present invention is an activating fusion gene.

In one embodiment of the present invention the fusion gene comprises one or more nucleotide

sequences encoding a protein or a fragment of the protein, said protein being selected from the group consisting of a guanine nucleotide exchange factor for RAS proteins, an adenylyl cyclase, a Zink finger protein, a protein tyrosine kinase, a member of ADP-ribosylation factor GTPase activating (ARF-GAP) protein family, a member of the IGF-1R pathway, a polypeptide hormone precursor, an adducin, a fibronectin type III domain containing protein, a germinal center kinase (GCKs), an chromosome open reading frame, a member of the Ras-association domain family (RASSF), a member of the E2F family of transcription factors, a ribosomal protein of the L14E family, a member of CDS1 subfamily of serine/threonine protein kinases, a THO complex, a member of P-type lectin family, a membrane glycoprotein of T-Lymphocytes, a Z-DNA binding protein, a breast carcinoma amplified sequence, an early B-cell factor, a class V ATPase, a protein involved in glycosylphosphatidylinositol (GPI)-anchor biosynthesis, a member of the ESCRT-III protein complex, a repressor of glucocorticoid receptor, a discs, large (Drosophila) homolog-associated protein, a G-protein coupled receptor, and a cholesterol ester hydrolase; and wherein the fusion gene may comprise one or more nucleotide sequences from a non-protein coding RNA, preferably a miscRNA, or a fragment thereof.

In a preferred embodiment the present invention the fusion gene comprises one or more nucleotide sequences encoding a protein or a fragment of the protein, said protein being selected from the group consisting of a guanine nucleotide exchange factor for RAS proteins, an adenylyl cyclase, a Zink finger protein, a protein tyrosine kinase, a fibronectin type III domain containing protein, and a germinal center kinase (GCKs), or a fragment thereof. Preferably, said fusion gene comprises one or more nucleotide sequence of a gene, or a fragment of the gene, said gene having a sequence with at least 70% identity to a sequence selected from the group consisting of SOS1 (SEQ ID NO. 1), ADCY3 (SEQ ID NO. 5), ZNF142 (SEQ ID NO. 11), PTK6 (SEQ ID NO. 15), FNDC3B (SEQ ID NO. 40), and TNIK (SEQ ID NO. 44). It is further preferred that the fusion gene comprises one or more nucleotide sequences of a fragment of the gene, said fragment having a sequence with at least 70% identity to a sequence selected from the group consisting of SOS1 (SEQ ID NO. 3), ADCY3 (SEQ ID NO. 7), ZNF142 (SEQ ID NO. 13), PTK6 (SEQ ID NO. 17), FNDC3B (SEQ ID NO. 42), and TNIK (SEQ ID NO. 46). In a particular preferred embodiment the fusion gene is selected from the group of fusion genes consisting of SOS1-ADCY3, ZNF142-PTK6, and FNDC3B-TNIK; preferably the fusion gene has a sequence with at least 70% identity to a sequence selected

from the group consisting of SOS1-ADCY3 (SEQ ID NO. 9), ZNF142-PTK6 (SEQ ID NO. 19), and FNDC3B-TNIK (SEQ ID NO. 48). The identity of the proteins encoded by the nucleotide sequence comprised in the fusion gene as outlined above is preferably at least 80% to the recited sequences, more preferably at least 85%, even more preferably at least 90%, yet
5 more preferably at least 95%, even more preferably at least 99%.

The inventors found several fusion genes comprising one or more of the above mentioned genes.

- 10 One preferred activating fusion gene comprises SOS1. SOS1 encodes guanine nucleotide exchange factor for RAS proteins, which are membrane proteins that bind guanine nucleotides and participate in signal transduction pathways. The product of this gene regulates RAS proteins by facilitating the exchange of GTP for GDP. Mutations in this gene are associated with gingival fibromatosis 1 and Noonan syndrome type 4. The fusion gene comprising SOS1
15 also comprised a fragment of ADCY3. ADCY3 encodes adenylyl cyclase 3, which is a membrane-associated enzyme and catalyzes the formation of the secondary messenger cyclic adenosine monophosphate (cAMP). This protein appears to be widely expressed in various human tissues and may be involved in a number of physiological and pathophysiological metabolic processes. The present inventors detected in the EML4-ALK-positive non-small cell
20 lung cancer (NSCLC) cell-line H3122, and through the next generation sequencing technique RNA-seq, the presence of a novel fusion polynucleotide of a partial SOS1 fused with a partial ADCY3 (hereinafter referred to as SOS1-ADCY3(SEQ ID NO. 9)), which is generated by an intra-chromosomal rearrangement in chromosome 2. The inventors successfully isolated the cDNA of SOS1-ADCY3 fusion gene and the whole ORF of the fusion gene was sequenced.
25 Based on the data here presented SOS1-ADCY3 appears to be a modulator of the activity of EML4-ALK and plays a role in the response to ALK inhibitors. Moreover, patients which tumors carry this fusion gene may benefit from MEK inhibitor therapy. Hence, subject-matter relating to an SOS1-ADCY3 fusion gene is particularly preferred. In a preferred embodiment the presence and/or expression of a SOS1-ADCY3 fusion gene is attributed to the presence of
30 adenocarcinoma. and FNDC3B-TNIK for adeno, ZNF142-PTK6 for large-cell neuroendocrine

A further activating fusion gene comprised a fragment of ZNF142. ZNF142 encodes for a zinc

finger protein, which are regulatory proteins, involved in many cellular functions. Zinc finger proteins bind primarily to DNA and in most cases they attach to regions near certain genes and turn the genes on and off as needed. The same fusion gene also comprised a fragment of the PTK6 gene. Protein tyrosine kinase 6 (PTK6), also referred to as breast tumour kinase BRK, is a member of a distinct family of kinases that is evolutionarily related to the SRC family of tyrosine kinases. Functions of the PTK6 family of intracellular tyrosine kinases are still, for the most part, poorly understood. PTK6 has been linked to differentiation, regulation of AKT, growth regulation and apoptosis and its ability to enter the nucleus sets them apart from the more intensely studied SRC family of kinases. In contrast to its functions in normal tissues, PTK6 appears to promote oncogenic signalling in epithelial tumour cells. PTK6 was found expressed in greater than 60% of breast tumours and breast cancer derived cell-lines, but was absent in normal mammary tissue and benign lesions. Moreover, the protein was over-expressed in 86% of invasive ductal breast tumours. In addition to this, amplification of the PTK6 gene was found to correlate with amplification of the ERBB2 locus in a panel of 202 human breast cancer samples. It has also been shown that cytoplasmic retention of PTK6 promotes growth of prostate tumour cells. Finally, PTK6 somatic mutations were found in 2.5% of human melanomas as well as in the RT-4 bladder and SW900 lung cancer cell-lines. The present inventors detected in a small cell lung cancer (SCLC) primary tumour, and through the next generation sequencing technique RNA-seq, the presence of a somatic novel fusion gene comprising a fragment of ZNF142 fused to a fragment of PTK6. The fusion detected was due to a genomic rearrangement between chromosome 2 and chromosome 20. The inventors successfully isolated the cDNA of ZNF142-PTK6 fusion gene and the whole ORF of the fusion gene was sequenced. Hence, in one embodiment of the present invention the presence and/or expression of a fusion gene comprising a nucleic acid sequence encoding a protein or fragment of the protein, said protein being a Zinc finger protein, preferably ZNF142, or a protein tyrosine kinase, preferably PTK6, is attributed to large-cell neuroendocrine tumor. In this context it has surprisingly been found that the fusion gene identified resulted in a new kinase domain sequence, which only in part is the same as the original kinase domain of PTK6. The reason for the creation of a new kinase domain instead of keeping the one of the PTK6 could be explained by the fact that the SH2 domain located upstream of the wild-type PTK6 kinase domain is required for both negative and positive regulation of the wild-type PTK6 kinase activity. Thus an oncogenic active kinase fusion could not have been created by simple fusion of the kinase

domain to any partner with a constitutive promoter. The novel kinase domain has the sequence of SEQ ID NO. 38. Hence, the present invention also relates to proteins comprising a sequence according to SEQ ID NO. 38 or a sequence at least 70% identical thereto, preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, yet more preferably at least 95%, even more preferably at least 99% identical. Furthermore the invention also relates to a nucleic acid sequence comprising a sequence encoding an amino acid sequence according to SEQ ID NO. 38 or a sequence at least 70% identical thereto, preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, yet more preferably at least 95%, even more preferably at least 99% identical. Furthermore, the invention also relates to nucleic acids encoding a protein comprising a sequence according to SEQ ID NO. 38 or a sequence at least 70% identical thereto, preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, yet more preferably at least 95%, even more preferably at least 99% identical. Furthermore, the invention also relates to a nucleic acid sequence according to SEQ ID NO. 150 or a sequence at least 70% identical thereto, preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, yet more preferably at least 95%, even more preferably at least 99% identical.

A further activating fusion gene associated with lung cancer found by the inventors comprised a fragment of the AGAP1 gene. The protein encoded by AGAP1 belongs to an ADP-ribosylation factor GTPase-activating (ARF-GAP) protein family involved in membrane traffic and actin cytoskeleton dynamics. A second fragment encompassed in said fusion gene is a fragment of the IGFBP2 gene which encodes one of the 6 proteins. These proteins which regulate the type I insulin-like growth factor receptor (IGF-1R) pathway ligands, IGF-1 and IGF-II. The IGF-1R pathway appears to play important roles in oncogenicity, metastasis, and resistance to existing forms of anti-cancer therapy. The present inventors detected in a large cell neuroendocrine lung cancer (LCNEC) primary tumour, and through the next generation sequencing technique RNA-seq, the presence of a novel fusion polynucleotide of a partial AGAP1 fused with a partial IGFBP2 (hereinafter referred to as AGAP1-IGFBP2 (SEQ ID NO. 29), which is generated by an intra-chromosomal rearrangement in chromosome 2. The inventors successfully isolated the cDNA of AGAP1-IGFBP2 fusion gene and the whole ORF was sequenced

A further preferred activating fusion gene comprised a fragment of the FNDC3B gene. FNDC3B is fibronectin type III domain containing 3B. Furthermore, a fragment of the TNIK gene is comprised. TNIK is a germinal centre kinases (GCKs) characterized by an N-terminal kinase domain and a C-terminal GCK domain that serves a regulatory function. The present
5 inventors detected in a lung AD primary tumour, and through the next generation sequencing technique RNA-seq, the presence of a novel fusion polynucleotide of a partial FNDC3B fused with a partial TNIK (hereinafter referred to as FNDC3B-TNIK (SEQ ID NO. 48)), which is generated by an intra-chromosomal rearrangement in chromosome 3. The inventors successfully isolated the cDNA of the FNDC3B-TNIK fusion gene and the whole ORF was
10 sequenced. TNIK is a member of the germinal center kinase (GCK) family of protein kinases, which have an N-terminal kinase domain and a C-terminal regulatory region (CNH). In the detected FNDC3B-TNIK fusion gene only the C-terminal regulatory domain is preserved. This regulatory domain is known to be necessary and sufficient for activation of the JNK pathway (the kinase domain is dispensable), which downstream target. Furthermore, TNIK seems to be
15 very low expressed in adenocarcinomas (2 FPKM on average). The deregulation of TNIK expression by the presently presented fusion gene results in an oncogenicity making the FNDC3B-TNIK fusion gene and the related subject-matter one of the preferred embodiments of the present invention.

20 The inventors could also identify an activating fusion gene comprising a fragment of C12orf11 which is chromosome 12 open reading frame 11. The fusion gene also comprised a fragment of the RASSF8 gene, the encoding a member of the Ras-association domain family (RASSF) of tumour suppressor proteins. This gene is essential for maintaining adherent junctions function in epithelial cells and has a role in epithelial cell migration. It is a lung tumour suppressor gene
25 candidate. Multiple alternatively spliced transcript variants encoding different isoforms have been found for this gene. The present inventors detected in a lung AD primary tumour, and through the next generation sequencing technique RNA-seq, the presence of a novel fusion polynucleotide of a partial C12orf11 fused with a partial RASSF8 (hereinafter referred to as C12orf11-RASSF8 (SEQ ID NO. 58)), which is generated by an intra-chromosomal
30 rearrangement in chromosome 12. The inventors successfully isolated the cDNA of the C12orf11-RASSF8 fusion gene and the whole ORF was sequenced.

The inventors found an inactivating fusion gene comprising a fragment of the CHECK2 gene. The protein encoded by CHEK2 is a cell cycle checkpoint regulator and putative tumor suppressor. It contains a forkhead-associated protein interaction domain essential for activation in response to DNA damage and is rapidly phosphorylated in response to replication blocks and DNA damage. When activated, the encoded protein is known to inhibit CDC25C phosphatase, preventing entry into mitosis, and has been shown to stabilize the tumor suppressor protein p53, leading to cell cycle arrest in G1. In addition, this protein interacts with and phosphorylates BRCA1, allowing BRCA1 to restore survival after DNA damage. Mutations in this gene have been linked with Li-Fraumeni syndrome, a highly penetrant familial cancer phenotype usually associated with inherited mutations in TP53. Also, mutations in this gene are thought to confer a predisposition to sarcomas, breast cancer, and brain tumors. This nuclear protein is a member of the CDS1 subfamily of serine/threonine protein kinases. Three transcript variants encoding different isoforms have been found for this gene. The fusion gene further comprised a fragment of the THOC5 (THO complex 5) gene. The present inventors detected in a lung AD primary tumour, and through the next generation sequencing technique RNA-seq, the presence of a novel fusion gene of a fragment of CHEK2 fused with a fragment of THOC5 (hereinafter referred to as CHEK2-THOC5 (SEQ ID NO. 77)), which is generated by an intra-chromosomal rearrangement in chromosome 22. This rearrangement leads to an early stop codon resulting in a predicted loss of function of both protein partners.

A further inactivating fusion gene comprised a fragment of the M6PR gene. M6PR encodes a member of the P-type lectin family. P-type lectins play a critical role in lysosome function through the specific transport of mannose-6-phosphate-containing acid hydrolases from the Golgi complex to lysosomes. The encoded protein functions as a homodimer and requires divalent cations for ligand binding. Alternatively spliced transcript variants encoding multiple isoforms have been observed for this gene. A pseudogene of this gene is located on the long arm of chromosome X. The fusion gene further comprised a fragment of the CD4 gene. CD4 encodes a membrane glycoprotein of T lymphocytes that interacts with major histocompatibility complex class II antigens and is also a receptor for the human immunodeficiency virus. This gene is expressed not only in T lymphocytes, but also in B cells, macrophages, and granulocytes. It is also expressed in specific regions of the brain. The protein functions to initiate or augment the early phase of T-cell activation, and may function as an

important mediator of indirect neuronal damage in infectious and immune-mediated diseases of the central nervous system. Multiple alternatively spliced transcript variants encoding different isoforms have been identified in this gene. The present inventors detected in a LCNEC primary tumour, and through the next generation sequencing technique RNA-seq, the presence of a novel fusion gene of a partial M6PR fused with a partial CD4 (hereinafter referred to as M6PR-CD4 (SEQ ID NO. 86)), which is generated by an intra-chromosomal rearrangement in chromosome 12. This rearrangement leads to an early stop codon resulting in a predicted loss of function of both protein partners.

One further identified rearrangement comprises a fragment of the ZBP1 gene. ZBP1 encodes a Z-DNA binding protein. Z-DNA formation is a dynamic process, largely controlled by the amount of supercoiling. Furthermore, the translocation comprised a fragment of the BCAS1 gene. BCAS1 resides in a region at 20q13, which is amplified in a variety of tumor types and associated with more aggressive tumor phenotypes. Among the genes identified from this region, it was found to be highly expressed in three amplified breast cancer cell lines and in one breast tumor without amplification at 20q13.2. However, this gene is not in the common region of maximal amplification and its expression was not detected in the breast cancer cell line MCF7, in which this region is highly amplified. The inventors detected in a lung AD primary tumour, and through the next generation sequencing technique RNA-seq, the presence of a novel rearrangement of a ZBP1 fragment fused with a BCAS1 fragment (hereinafter referred to as ZBP1-BCAS1 (SEQ ID NO. 95)), which is generated by an intra-chromosomal rearrangement in chromosome 20. This rearrangement leads to an early stop codon resulting in a predicted loss of function of both protein partners.

The inventors furthermore detected in a LCNEC primary tumour, and through the next generation sequencing technique RNA-seq, the presence of a novel fusion polynucleotide of an EBF1 fragment fused with the 5' UTR region of the ATP10B gene (hereinafter referred to as EBF1-ATP10B (SEQ ID NO. 104)), which is generated by an intra-chromosomal rearrangement in chromosome 5. EBF1 is early B-cell factor 1. ATP10B is ATPase, class V, type 10B. This rearrangement leads to an early stop codon resulting in a predicted loss of function of both protein partners.

A further inactivating fusion gene comprised a fragment of PIGF. PIGF encodes a protein involved in glycosylphosphatidylinositol (GPI)-anchor biosynthesis. The GPI-anchor, a glycolipid containing three mannose molecules in its core backbone, is found on many blood cells where it serves to anchor proteins to the cell surface. The encoded protein and another GPI synthesis protein, PIGO, function in the transfer of ethanolaminephosphate to the third mannose in GPI. Alternatively spliced transcript variants encoding different isoforms have been described. Furthermore this fusion gene comprised a fragment of the *CHMP3* gene. *CHMP3* encodes a protein that sorts transmembrane proteins into lysosomes/vacuoles via the multivesicular body (MVB) pathway. This protein, along with other soluble coiled-coil containing proteins, forms part of the ESCRT-III protein complex that binds to the endosomal membrane and recruits additional cofactors for protein sorting into the MVB. This protein may also co-immunoprecipitate with a member of the IFG-binding protein superfamily. Alternative splicing results in multiple transcript variants. Read-through transcription also exists between this gene and the upstream ring finger protein 103 (RNF103) gene. The inventors detected in a lung AD primary tumour, and through the next generation sequencing technique RNA-seq, the presence of a novel fusion gene comprising the 5' UTR region of PIGF fused with a fragment of *CHMP3* (hereinafter referred to as PIGF- *CHMP3* (SEQ ID NO. 113), which is generated by an intra-chromosomal rearrangement in chromosome 2. This rearrangement leads to an early stop codon resulting in a predicted loss of function of both protein partners.

A further identified inactivating fusion gene comprises a fragment of the GRLF1 gene (human glucocorticoid receptor DNA binding factor), which associates with the promoter region of the glucocorticoid receptor gene (hGR gene), is a repressor of glucocorticoid receptor transcription. The amino acid sequence deduced from the cDNA sequences show the presence of three sequence motifs characteristic of a zinc finger and one motif suggestive of a leucine zipper in which 1 cysteine is found instead of all leucines. The GRLF1 enhances the homologous down-regulation of wild-type hGR gene expression. Biochemical analysis suggests that GRLF1 interaction is sequence specific and that transcriptional efficacy of GRLF1 is regulated through its interaction with specific sequence motif. The level of expression is regulated by glucocorticoids. A fragment of DLGAP1 (discs, large (Drosophila) homolog-associated protein 1) is also comprised in said fusion gene. The inventors detected in a FGFR1-amplified SQ primary tumor, and through the next generation sequencing technique RNA-seq, the

presence of a novel fusion gene comprising a fragment of GRLF1 fused to a DLGAP1 fragment (hereinafter referred to as GRLF1-DLGAP1 (SEQ ID NO. 122), which is generated by a genomic rearrangement between chromosome 18 and chromosome 19. This rearrangement leads to an early stop codon resulting in a predicted loss of function of both protein partners.

5

An additional inactivating fusion gene identified comprises a fragment of the E2F3 gene. The protein encoded by this gene is a member of the E2F family of transcription factors. The E2F family plays a crucial role in the control of cell cycle and action of tumor suppressor proteins and is also a target of the transforming proteins of small DNA tumor viruses. The E2F proteins contain several evolutionally conserved domains found in most members of the family. These domains include a DNA binding domain, a dimerization domain which determines interaction with the differentiation regulated transcription factor proteins (DP), a transactivation domain enriched in acidic amino acids, and a tumor suppressor protein association domain which is embedded within the transactivation domain. This protein and another 2 members, E2F1 and E2F2, have an additional cyclin binding domain. This protein binds specifically to retinoblastoma protein pRB in a cell-cycle dependent manner. Two transcript variants encoding different isoforms have been found for this gene. The inactivating fusion gene further comprises a fragment of FLJ22536 which is long intergenic non-protein coding RNA 340. The present inventors detected in a FGFR1-amplified SQ primary tumour, and through the next generation sequencing technique RNA-seq, the presence of a novel fusion gene of a fragment of E2F3 fused with a fragment of long intergenic non-protein coding RNA 340 (hereinafter referred to as E2F3-FLJ22536), which is generated by an intra-chromosomal rearrangement in chromosome 6. This rearrangement leads to an early stop codon resulting in a predicted loss of function of E2F3.

25

A further inactivating fusion gene comprises a fragment of the GPR160 gene and a fragment of the NCEH1 gene. GPR160 is G protein-coupled receptor 160. NCEH1 is neutral cholesterol ester hydrolase 1. The inventors detected in a SCLC primary tumour, and through the next generation sequencing technique RNA-seq, the presence of a novel fusion gene, comprising the of the 5' UTR region of the GPR160 gene fused with a fragment of the NCEH1 gene (hereinafter referred to as GPR160-NCEH1(SEQ ID NO. 139)), which is generated by an intra-chromosomal rearrangement in chromosome 3. This rearrangement leads to an early stop codon

30

resulting in a predicted loss of function of both protein partners.

A transcription-mediated fusion gene found comprised a fragment of the *POMC* gene, which encodes a polypeptide hormone precursor that undergoes extensive, tissue-specific, post-translational processing via cleavage by subtilisin-like enzymes known as prohormone convertases. *POMC* is the precursor of ACTH (adrenocorticotrophic) hormone, which is a hormone produced and secreted by the anterior pituitary gland. It is an important component of the hypothalamic-pituitary-adrenal axis and is often produced in response to biological stress. Its principal effects are increased production and release of corticosteroids and cortisol from the adrenal cortex. A tumour (adenoma) in the pituitary gland can produce large amounts of ACTH, which in turn elevates cortisol in the blood leading to the hormone disorder Cushing's syndrome. Also, tumours outside the normal pituitary-adrenal system can produce ACTH that affects the adrenal glands. This aetiology is called ectopic or paraneoplastic Cushing's syndrome and is seen in diseases like small cell lung cancer. This transcription-mediated fusion gene further comprised a fragment of *ADD1*. *ADD1* belongs to the family of adducins, which is a family of cytoskeleton proteins encoded by three genes (alpha, beta, gamma). Adducin binds with high affinity to Ca^{2+} /calmodulin and is a substrate for protein kinases A and C. The present inventors detected in two lung carcinoid (CA) primary tumours, and through the next generation sequencing technique RNA-seq, the presence of a novel rearrangement between the 3'UTR of *POMC* with the last part of the 3'UTR of *ADD1*. Since the chimeric transcript was only detected at RNA level but not at genomic level, this suggests that this chimeric transcript is generated by trans-splicing.

One regulatory fusion gene found by the inventors comprised a fragment of the gene *E2F4*. *E2F4* belongs to E2F family of transcription factors which plays a crucial role in the control of cell cycle and action of tumour suppressor proteins and is also a target of the transforming proteins of small DNA tumour viruses. The E2F proteins contain several evolutionally conserved domains found in most members of the family. These domains include a DNA binding domain, a dimerization domain which determines interaction with the differentiation regulated transcription factor proteins (DP), a transactivation domain enriched in acidic amino acids, and a tumour suppressor protein association domain which is embedded within the transactivation domain. This protein binds to all three of the tumour suppressor proteins pRB,

p107 and p130, but with higher affinity to the last two. It plays a role in the suppression of proliferation-associated genes, and its gene mutation and increased expression may be associated with human cancer. In this fusion gene an *E2F4* fragment was fused to a fragment of the *RPL14* gene. *RPL14* encodes a ribosomal protein that is a component of the 60S subunit.

5 Ribosomes, the organelles that catalyse protein synthesis, consist of a small 40S subunit and a large 60S subunit. Together these subunits are composed of 4 RNA species and approximately 80 structurally distinct proteins. The protein belongs to the L14E family of ribosomal proteins. It contains a basic region-leucine zipper (bZIP)-like domain. The protein is located in the cytoplasm. This gene contains a trinucleotide (GCT) repeat tract whose length is highly
10 polymorphic; this triplet repeats result in a stretch of alanine residues in the encoded protein. Transcript variants utilizing alternative polyA signals and alternative 5'-terminal exons exist but all encode the same protein. As is typical for genes encoding ribosomal proteins, there are multiple processed pseudo genes of this gene dispersed through the genome. The inventors detected in a FGFR1-amplified NSCLC cell-line and in a FGFR1-amplified SQ lung cancer
15 primary tumour, and through the next generation sequencing technique RNA-seq, the presence of a novel fusion polynucleotide of a partial *E2F4* fused with a partial *RPL14* (hereinafter referred to as *E2F4-RPL14* (SEQ ID NO. 68). The fusion gene was only detected at RNA level but not at genomic level, this shows that this fusion gene is generated by trans-splicing.

20 In one embodiment said guanine nucleotide exchange factor for RAS proteins is SOS1 (SEQ ID NO. 2) or a protein having at least 70% identity thereto, preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, yet more preferably at least 95%, even more preferably at least 99% identical; said adenylyl cyclase is selected from the group of adenylyl cyclases consisting of ADCY1 (SEQ ID NO. 141), ADCY2 (SEQ ID NO. 142),
25 ADCY3 (SEQ ID NO. 6), ADCY4 (SEQ ID NO. 143), ADCY5 (SEQ ID NO. 144), ADCY6 (SEQ ID NO. 145), ADCY7 (SEQ ID NO. 146), ADCY8 (SEQ ID NO. 147), ADCY9 (SEQ ID NO. 148), and ADCY10 (SEQ ID NO. 149); or a protein having at least 70% identity thereto; said Zink finger protein is ZNF142 (SEQ ID NO. 12) or a protein having at least 70% identity thereto; said protein tyrosine kinase is PTK6 (SEQ ID NO. 16) or a protein having at
30 least 70% identity thereto; said member of ADP-ribosylation factor GTPase activating (ARF-GAP) protein family is AGAP1 (SEQ ID NO. 22) or a protein having at least 70% identity thereto; said member of the IGF-1R pathway is IGFBP2 (SEQ ID NO. 26) or a protein having

at least 70% identity thereto; said polypeptide hormone precursor is POMC (SEQ ID NO. 32) or a protein having at least 70% identity thereto; said adducin is ADD1 (SEQ ID NO. 36) or a protein having at least 70% identity thereto; said fibronectin type III domain containing protein is FNDC3B (SEQ ID NO. 41) or a protein having at least 70% identity thereto; said germinal center kinase (GCK) is TNIK (SEQ ID NO. 45) or a protein having at least 70% identity thereto; said chromosome open reading frame is C12orf11 (SEQ ID NO. 51) or a protein having at least 70% identity thereto; said member of the Ras-association domain family (RASSF) is RASSF8 (SEQ ID NO. 55) or a protein having at least 70% identity thereto; said member of the E2F family of transcription factors is selected from the group consisting of E2F4 (SEQ ID NO. 61) and E2F3 (SEQ ID NO. 125), or a protein having at least 70% identity thereto; said ribosomal protein of the L14E family is RPL14 (SEQ ID NO. 65) or a protein having at least 70% identity thereto; said member of CDS1 subfamily of serine/threonine protein kinases is CHEK2 (SEQ ID NO. 71) or a protein having at least 70% identity thereto; said THO complex is THOC5 (SEQ ID NO. 75) or a protein having at least 70% identity thereto; said member of P-type lectin family is M6PR (SEQ ID NO. 80) or a protein having at least 70% identity thereto; said membrane glycoprotein of T-Lymphocytes is CD4 (SEQ ID NO. 84) or a protein having at least 70% identity thereto; said a Z-DNA binding protein is ZBP1 (SEQ ID NO. 89) or a protein having at least 70% identity thereto; said breast carcinoma amplified sequence is BCAS1 (SEQ ID NO. 93) or a protein having at least 70% identity thereto; said early B-cell factor is EBF1 (SEQ ID NO. 98) or a protein having at least 70% identity thereto; said class V ATPase is ATP10B (SEQ ID NO. 102) or a protein having at least 70% identity thereto; said protein involved in glycosylphosphatidylinositol (GPI)-anchor biosynthesis is PIGF (SEQ ID NO. 107) or a protein having at least 70% identity thereto; said member of the ESCRT-III protein complex is CHMP3 (SEQ ID NO. 111) or a protein having at least 70% identity thereto; said repressor of glucocorticoid receptor is GRLF1 (SEQ ID NO. 116) or a protein having at least 70% identity thereto; said discs, large (*Drosophila*) homolog-associated protein is DLGAP1 (SEQ ID NO. 120) or a protein having at least 70% identity thereto; said G-protein coupled receptor is GPR160 (SEQ ID NO. 133) or a protein having at least 70% identity thereto; said cholesterol ester hydrolase is NCEH1 (SEQ ID NO. 137) or a protein having at least 70% identity thereto; and said non-protein coding RNA is FLJ22536 (SEQ ID. 128). The identity of the proteins encoded by the nucleotide sequence comprised in the fusion gene as outlined above is preferably at least 80%, more preferably at least 85%, even

more preferably at least 90%, yet more preferably at least 95%, even more preferably at least 99%. It is also preferred that the proteins encoded by the nucleotide sequence comprised in the fusion gene as outlined above are identical to the sequences outlined above.

5 The inventors found that fusion genes comprising a guanine nucleotide exchange factor for RAS proteins are oncogenic and highly predictive for lung cancer. One preferred guanine nucleotide exchange factor for RAS proteins being of particular predictive value is SOS1. SOS1 (Gene ID: 6654) encodes guanine nucleotide exchange factor for RAS proteins, which are membrane proteins that bind guanine nucleotides and participate in signal transduction
10 pathways. The product of this gene may regulate RAS proteins by facilitating the exchange of GTP for GDP. Mutations in this gene are associated with gingival fibromatosis 1 and Noonan syndrome type 4. In a preferred embodiment said guanine nucleotide exchange factor for RAS proteins is SOS1 (SEQ ID NO. 2) or a protein having at least 70% identity thereto, preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%,
15 yet more preferably at least 99% identity. In a further preferred embodiment said guanine nucleotide exchange factor for RAS proteins is SOS1 (SEQ ID NO. 2) and has the sequence of SEQ ID NO. 2.

It will be readily understood that the nucleotide sequence encoding guanine nucleotide
20 exchange factor for RAS proteins may have any sequence as long as the nucleotide sequence comprises a sequence or fragment encoding an outlined protein sequence. However, in a preferred embodiment the nucleotide sequence encoding a guanine nucleotide exchange factor for RAS proteins comprises the sequence of the SOS1 gene or a fragment thereof. In one embodiment the nucleotide sequence encoding a guanine nucleotide exchange factor for RAS
25 proteins comprises the sequence of SOS1 (SEQ ID NO. 1). Preferably the nucleotide sequence encoding a fragment of a guanine nucleotide exchange factor for RAS proteins comprises a fragment of the sequence of SOS1 (SEQ ID NO. 1). A preferred nucleotide sequence encoding a fragment of a guanine nucleotide exchange factor comprises a sequence encoding for a fragment of SOS1 comprising the sequence of SEQ ID NO. 4 or a sequence having 70%,
30 preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99%.. In yet a further embodiment the nucleotide sequence encoding a fragment of a guanine nucleotide exchange factor comprises a sequence

encoding a protein identical to the sequence of SEQ ID NO. 4. In a preferred embodiment the fusion gene comprises a fragment of SOS1 (SEQ ID NO. 1). The fragment of SOS1 (SEQ ID NO. 1) in one embodiment comprises the sequence of SEQ ID NO. 3. Hence, in one embodiment the fusion gene comprises a nucleotide sequence according to SEQ ID NO. 3 or a sequence having at least 70%, preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99%. In yet a further embodiment the fusion gene comprises a nucleotide sequence identical to SEQ ID NO. 3.

In a further embodiment the fusion gene comprises a nucleotide sequence encoding a protein or fragment thereof, said protein being an adenylyl cyclase. Adenylyl cyclases are known by those of ordinary skills and may be of any type. However, in a preferred embodiment the adenylyl cyclase is selected from the group of adenylyl cyclases consisting of ADCY1 (SEQ ID NO. 141), ADCY2 (SEQ ID NO. 142), ADCY3 (SEQ ID NO. 6), ADCY4 (SEQ ID NO. 143), ADCY5 (SEQ ID NO. 144), ADCY6 (SEQ ID NO. 145), ADCY7 (SEQ ID NO. 146), ADCY8 (SEQ ID NO. 147), ADCY9 (SEQ ID NO. 148), and ADCY10 (SEQ ID NO. 149); or a protein having at least 70% identity thereto, preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99%. In a further preferred embodiment said adenylyl cyclase has the sequence of any one of the adenylyl cyclases selected from the group consisting of ADCY1 (SEQ ID NO. 141), ADCY2 (SEQ ID NO. 142), ADCY3 (SEQ ID NO. 6), ADCY4 (SEQ ID NO. 143), ADCY5 (SEQ ID NO. 144), ADCY6 (SEQ ID NO. 145), ADCY7 (SEQ ID NO. 146), ADCY8 (SEQ ID NO. 147), ADCY9 (SEQ ID NO. 148), and ADCY10 (SEQ ID NO. 149). The inventors unexpectedly found that a fusion gene comprising a nucleotide sequence coding for ADCY3 or a fragment is especially well suited for the present invention. The ADCY3-gene (Gene ID: 109) encodes adenylyl cyclase 3, which is a membrane-associated enzyme and catalyzes the formation of the secondary messenger cyclic adenosine monophosphate (cAMP). This protein is widely expressed in various human tissues and involved in a number of physiological and pathophysiological metabolic processes. Hence, in a preferred embodiment of the present invention the adenylyl cyclase is ADCY3 (SEQ ID NO. 6) or a protein having at least 70% identity thereto, preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99%. In a further preferred embodiment said adenylyl cyclase has the sequence of ADCY3 (SEQ ID NO. 6). However, in a

preferred embodiment the nucleotide sequence encoding an adenylyl cyclase comprises the sequence of the ADCY3 gene or a fragment thereof. In one embodiment the nucleotide sequence encoding an adenylyl cyclase comprises the sequence of ADCY3 (SEQ ID NO. 5). Preferably the nucleotide sequence encoding a fragment of an adenylyl cyclase comprises a fragment of the sequence of ADCY3 (SEQ ID NO. 5). A preferred nucleotide sequence encoding a fragment of an adenylyl cyclase comprises a sequence encoding for a fragment of ADCY3 comprising the sequence of SEQ ID NO. 8 or a sequence having 70%, preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99%. In yet a further embodiment the nucleotide sequence encoding a fragment of an adenylyl cyclase comprises a sequence encoding a protein identical to the sequence of SEQ ID NO. 8. In a preferred embodiment the fusion gene comprises a fragment of the ADCY3 gene (SEQ ID NO. 5). The fragment of the ADCY3 gene (SEQ ID NO. 5) in one embodiment comprises the sequence of SEQ ID NO. 7. Hence, in one embodiment the fusion gene comprises a nucleotide sequence according to SEQ ID NO. 7 or a sequence having at least 70%, preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99%. In yet a further embodiment the fusion gene comprises a nucleotide sequence identical to SEQ ID NO. 7.

In one embodiment the fusion gene comprises a nucleotide sequence encoding a protein or fragment thereof, wherein said protein is a Zink finger protein. Preferably said Zink finger protein is ZNF142 (SEQ ID NO. 12) or a protein having at least 70% identity thereto; preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99%. In a further preferred embodiment said Zink finger protein has the sequence of ZNF142 (SEQ ID NO. 12). However, in a preferred embodiment the nucleotide sequence encoding a guanine nucleotide exchange factor for RAS proteins comprises the sequence of the ZNF142 gene or a fragment thereof. The ZNF142 gene (Gene ID: 7701) encodes for a zinc finger protein, which are regulatory proteins, involved in many cellular functions. Zinc finger proteins bind primarily to DNA and in most cases they attach to regions near certain genes and turn the genes on and off as needed. In one embodiment the nucleotide sequence encoding a Zink finger protein comprises the sequence of ZNF142 (SEQ ID NO. 11) or a fragment thereof, or a sequence having at least 70% identity thereto preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%,

yet more preferably at least 99%. Preferably the nucleotide sequence encoding a fragment of a Zink finger protein comprises a fragment of the sequence of ZNF142 (SEQ ID NO. 11). A preferred nucleotide sequence encoding a fragment of a Zink finger protein comprises a sequence encoding for a fragment of ZNF142 (SEQ ID NO. 11) comprising the sequence of
5 SEQ ID NO. 13 or a sequence having at least 70% identity thereto, preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99%. In yet a further embodiment the nucleotide sequence encoding a fragment of a Zink finger protein comprises a sequence encoding a protein with a identical to the sequence of SEQ ID NO. 14 or a sequence having at least 70% identity thereto, preferably
10 at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99%.

In one embodiment the fusion gene comprises a nucleotide sequence encoding a protein or fragment thereof, wherein said protein is a protein tyrosine kinase. Preferably said protein
15 tyrosine kinase is PTK6 (SEQ ID NO. 16) or a protein having at least 70% identity thereto; preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99%. In a further preferred embodiment said Zink finger protein has the sequence of PTK6 (SEQ ID NO. 16). However, in a preferred embodiment the nucleotide sequence encoding a protein tyrosine kinase comprises the sequence of the PTK6
20 gene or a fragment thereof. The PTK6 gene (Gene ID: 5753) also referred to as breast tumour kinase BRK, is a member of a distinct family of kinases that is evolutionarily related to the SRC family of tyrosine kinases. Functions of the PTK6 family of intracellular tyrosine kinases are still poorly understood. PTK6 has been linked to differentiation, regulation of AKT, growth regulation and apoptosis and its ability to enter the nucleus sets them apart from the more
25 intensely studied SRC family of kinases. In contrast to its functions in normal tissues, PTK6 appears to promote oncogenic signalling in epithelial tumour cells. PTK6 was found expressed in greater than 60% of breast tumours and breast cancer derived cell-lines, but was absent in normal mammary tissue and benign lesions. Moreover, the protein was over-expressed in 86% of invasive ductal breast tumours. In addition to this, amplification of the PTK6 gene was
30 found to correlate with amplification of the ERBB2 locus in a panel of 202 human breast cancer samples. It has also been shown that cytoplasmic retention of PTK6 promotes growth of prostate tumour cells. Finally, PTK6 somatic mutations were found in 2.5% of human

melanomas as well as in the RT-4 bladder and SW900 lung cancer cell-lines. However, the association of PTK6 comprised in fusion genes and lung cancer is not known so far. In one embodiment the nucleotide sequence encoding a protein tyrosine kinase comprises the sequence of PTK6 (SEQ ID NO. 15) or a fragment thereof, or a sequence having at least 70% identity thereto, preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99%. Preferably the nucleotide sequence encoding a fragment of a protein tyrosine kinase comprises a fragment of the sequence of PTK6 (SEQ ID NO. 15). A preferred nucleotide sequence encoding a fragment of a protein tyrosine kinase comprises a sequence encoding for a fragment of PTK6 (SEQ ID NO. 18), preferably the nucleotide sequence comprises the sequence of SEQ ID NO. 17 or a sequence having at least 70% identity thereto. In yet a further embodiment the nucleotide sequence encoding a fragment of a protein tyrosine kinase comprises a sequence encoding a protein identical to the sequence of SEQ ID NO. 18 or a sequence having at least 70% identity thereto, preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99%. In a further embodiment the fusion gene comprises a sequence encoding the novel kinase domain according to SEQ ID NO. 38 or a sequence having at least 70% identity thereto, preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99%.

In one embodiment the fusion gene comprises one or more nucleotide sequence of a gene or non-protein coding RNA; or a fragment of the gene or non-protein coding RNA, said gene or non-coding RNA having a sequence with at least 70% identity to a sequence selected from the group consisting of SOS1 (SEQ ID NO. 1), ADCY3 (SEQ ID NO. 5), ZNF142 (SEQ ID NO. 11), PTK6 (SEQ ID NO. 15), AGAP1 (SEQ ID NO. 21), IGFBP2 (SEQ ID NO. 25), POMC (SEQ ID NO. 31), ADD1 (SEQ ID NO. 35), FNDC3B (SEQ ID NO. 40), TNIK (SEQ ID NO. 44), C12orf11 (SEQ ID NO. 50), RASSF8 (SEQ ID NO. 54), E2F4 (SEQ ID NO. 60), RPL14 (SEQ ID NO. 64), CHEK2 (SEQ ID NO. 70), THOC5 (SEQ ID NO. 74), M6PR (SEQ ID NO. 79), CD4 (SEQ ID NO. 83), ZBP1 (SEQ ID NO. 88), BCAS1 (SEQ ID NO. 92), EBF1 (SEQ ID NO. 97), ATP10B (SEQ ID NO. 101), PIGF (SEQ ID NO. 106), CHMP3 (SEQ ID NO. 110), GRLF1 (SEQ ID NO. 115), DLGAP1 (SEQ ID NO. 119), E2F3 (SEQ ID NO. 124), GPR160 (SEQ ID NO. 132), NCEH1 (SEQ ID NO. 136); and FLJ22536 (SEQ ID NO. 128).

As outlined above the fusion gene according to the present invention may comprise only fragments of the genes. Preferably the fusion gene comprises one or more nucleotide sequences of a fragment of the gene or the non-protein coding RNA, said fragment having a sequence with at least 70% identity to a sequence selected from the group consisting of SOS1 (SEQ ID NO. 3), ADCY3 (SEQ ID NO. 7), ZNF142 (SEQ ID NO. 13), PTK6 (SEQ ID NO. 17), AGAP1 (SEQ ID NO. 23), IGFBP2 (SEQ ID NO. 27), POMC (SEQ ID NO. 33), ADD1 (SEQ ID NO. 37), FNDC3B (SEQ ID NO. 42), TNIK (SEQ ID NO. 46), C12orf11 (SEQ ID NO. 52), RASSF8 (SEQ ID NO. 56), E2F4 (SEQ ID NO. 62), RPL14 (SEQ ID NO. 66), CHEK2 (SEQ ID NO. 72), THOC5 (SEQ ID NO. 76), M6PR (SEQ ID NO. 81), CD4 (SEQ ID NO. 85), ZBP1 (SEQ ID NO. 90), BCAS1 (SEQ ID NO. 94), EBF1 (SEQ ID NO. 99), ATP10B (SEQ ID NO. 103), PIGF (SEQ ID NO. 108), CHMP3 (SEQ ID NO. 112), GRLF1 (SEQ ID NO. 117), DLGAP1 (SEQ ID NO. 121), E2F3 (SEQ ID NO. 126), FLJ22536 (SEQ ID NO. 129), GPR160 (SEQ ID NO. 134), and NCEH1 (SEQ ID NO. 138). In a preferred embodiment the fusion gene comprises one or more nucleotide sequences of a fragment of the gene, said fragment having a sequence of at least 70% identity to a sequence selected from the group consisting of SOS1 (SEQ ID NO. 3), ADCY3 (SEQ ID NO. 7), ZNF142 (SEQ ID NO. 13), and PTK6 (SEQ ID NO. 17).

The fusion gene may comprise any of the above mentioned genes or non protein coding RNAs or fragments in any combination. However, the fusion gene preferably comprises just two of the above mentioned genes or non protein coding RNAs. Preferred combinations are selected from the group consisting of SOS1-ADCY3, ZNF142-PTK6, AGAP1-IGFBP2, POMC-ADD1, FNDC3B-TNIK, C12orf11-RASSF8, E2F4-RPL14, CHEK2-THOC5, M6PR-CD4, ZBP1-BCAS1, EBF1-ATP10B, PIGF-CHMP3, GRLF1-DLGAP1, E2F3-FLJ22536, and GPR160-NCEH1. Hence in a preferred embodiment the fusion gene is selected from the group of fusion genes consisting of SOS1-ADCY3, ZNF142-PTK6, AGAP1-IGFBP2, POMC-ADD1, FNDC3B-TNIK, C12orf11-RASSF8, E2F4-RPL14, CHEK2-THOC5, M6PR-CD4, ZBP1-BCAS1, EBF1-ATP10B, PIGF-CHMP3, GRLF1-DLGAP1, E2F3-FLJ22536, and GPR160-NCEH1. It is particularly preferred that the fusion gene is selected from the group of fusion genes consisting of SOS1-ADCY3, ZNF142-PTK6, and FNDC3B-TNIK.

As outlined above, the protein originally encoded by the comprised nucleotide sequences in the

fusion genes is not necessarily encoded by the fusion gene itself. However, in a preferred embodiment the at least a fragment of the protein originally encoded by the nucleotide sequence of at least one fusion partner is encoded by the fusion protein in its expressed form. In a preferred embodiment the fusion gene encodes a protein comprising a fragment selected from the group consisting of SOS1 (SEQ ID NO. 4), ADCY3 (SEQ ID NO. 8), ZNF142 (SEQ ID NO. 14), PTK6 (SEQ ID NO. 18), AGAP1 (SEQ ID NO. 24), IGFBP2 (SEQ ID NO. 28), POMC (SEQ ID NO. 34), FNDC3B (SEQ ID NO. 43), TNIK (SEQ ID NO. 47), C12orf11 (SEQ ID NO. 53), RASSF8 (SEQ ID NO. 57), E2F4 (SEQ ID NO. 63), RPL14 (SEQ ID NO. 67), CHEK2 (SEQ ID NO. 73), M6PR (SEQ ID NO. 82), ZBP1 (SEQ ID NO. 91), EBF1 (SEQ ID NO. 100), GRLF1 (SEQ ID NO. 118), and E2F3 (SEQ ID NO. 127). In a more preferred embodiment the fusion gene encodes a protein comprising a fragment selected from the group consisting of SOS1 (SEQ ID NO. 4), ADCY3 (SEQ ID NO. 8), ZNF142 (SEQ ID NO. 14), PTK6 (SEQ ID NO. 18), FNDC3B (SEQ ID NO. 43), and TNIK (SEQ ID NO. 47) or a sequence having at least 70% identity thereto, preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99%.

In a preferred embodiment the fusion gene has a sequence with at least 70% identity to a sequence selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 9), ZNF142-PTK6 (SEQ ID NO. 19), AGAP1-IGFBP2 (SEQ ID NO. 29), POMC-ADD1 (SEQ ID NO. 39), FNDC3B-TNIK (SEQ ID NO. 48), C12orf11-RASSF8 (SEQ ID NO. 58), E2F4-RPL14 (SEQ ID NO. 68), CHEK2-THOC5 (SEQ ID NO. 77), M6PR-CD4 (SEQ ID NO. 86), ZBP1-BCAS1 (SEQ ID NO. 95), EBF1-ATP10B (SEQ ID NO. 104), PIGF-CHMP3 (SEQ ID NO. 113), GRLF1-DLGAP1 (SEQ ID NO. 122), E2F3-FLJ22536 (SEQ ID NO. 130), and GPR160-NCEH1 (SEQ ID NO. 139).

In a preferred embodiment the fusion gene comprises two nucleotide sequences each encoding a protein or a fragment thereof, wherein the proteins are a guanine nucleotide exchange factor for RAS proteins and an adenylyl cyclase. In a further preferred embodiment the fusion gene comprises a nucleotide sequence encoding SOS1 or a fragment thereof and a nucleotide sequence encoding an adenylyl cyclase or a fragment thereof. In a further embodiment the fusion gene comprises a nucleotide sequence encoding a guanine nucleotide exchange factor for RAS proteins and a nucleotide sequence encoding a adenylyl cyclase selected from the

group consisting of ADCY1 (SEQ ID NO. 141), ADCY2 (SEQ ID NO. 142), ADCY3 (SEQ ID NO. 6), ADCY4 (SEQ ID NO. 143), ADCY5 (SEQ ID NO. 144), ADCY6 (SEQ ID NO. 145), ADCY7 (SEQ ID NO. 146), ADCY8 (SEQ ID NO. 147), ADCY9 (SEQ ID NO. 148), and ADCY10 (SEQ ID NO. 149), or a protein having at least 70% identity thereto. In yet a further preferred embodiment the fusion gene according to the invention comprises a nucleotide sequence encoding SOS1 (SEQ ID NO. 2) or a fragment thereof, or a protein having at least 70% identity thereto, preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably 99%, and a nucleotide sequence encoding adenylyl cyclase selected from the group consisting of ADCY1 (SEQ ID NO. 141), ADCY2 (SEQ ID NO. 142), ADCY3 (SEQ ID NO. 6), ADCY4 (SEQ ID NO. 143), ADCY5 (SEQ ID NO. 144), ADCY6 (SEQ ID NO. 145), ADCY7 (SEQ ID NO. 146), ADCY8 (SEQ ID NO. 147), ADCY9 (SEQ ID NO. 148), and ADCY10 (SEQ ID NO. 149), or a fragment thereof, or a protein having at least 70% identity thereto a sequence having at least 70% identity thereto, preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99%.

In an even more preferred embodiment the fusion gene is a SOS1-ADCY3 fusion gene. Preferably the SOS1-ADCY3 fusion gene comprises a nucleotide sequence encoding SOS1 (SEQ ID NO. 2), or a fragment thereof, or a protein having at least 70% identity thereto, preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99%, and a nucleotide sequence encoding ADCY3 (SEQ ID NO. 6) or a fragment thereof, or a protein having at least 70% identity thereto, preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99%. In a further preferred embodiment the SOS1-ADCY3 fusion gene comprises a nucleotide sequence encoding fragment of SOS1 (SEQ ID NO. 2) or a protein having at least 70% identity thereto, preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99% and nucleotide sequence encoding a fragment of ADCY3 (SEQ ID NO. 6) or a protein having at least 70% identity thereto, preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, yet more preferably at least 95%, even more preferably at least 99% identity. Preferred fragments are described herein above. In a preferred embodiment the fusion gene encodes the fusion protein comprising or consisting of SOS1-ADCY3 (SEQ ID NO. 10)

or a protein having at least 70% identity thereto, preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, yet more preferably at least 95%, even more preferably at least 99%. A preferred fusion gene comprises the nucleotide sequence of SOS1-ADCY3 (SEQ ID NO. 9) or a nucleotide sequence having at least 70% identity thereto, in a preferred embodiment the fusion gene has the nucleotide sequence of SOS1-ADCY3 (SEQ ID NO. 9), or at least 70% identity thereto, preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, yet more preferably at least 95%, even more preferably at least 99% identity.

10 In a further preferred embodiment the fusion gene comprises two nucleotide sequences each encoding a protein or a fragment thereof, wherein the proteins are a Zink finger protein and a protein tyrosine kinase. In a further preferred embodiment the fusion gene comprises a nucleotide sequence encoding ZNF142; or a fragment thereof; preferably encoding ZNF142 (SEQ ID NO. 12) or a fragment thereof, or a protein having at least 70% identity thereto, 15 preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, yet more preferably at least 95%, even more preferably at least 99% identity and a nucleotide sequence encoding a protein tyrosine kinase or a fragment thereof.

In a further embodiment the fusion gene comprises a nucleotide sequence encoding a Zink 20 finger protein and a nucleotide sequence encoding a protein tyrosine kinase 6 (PTK6), or a fragment thereof; preferably encoding PTK6 (SEQ ID NO. 16) or a fragment thereof, or a protein having at least 70% identity thereto, preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, yet more preferably at least 95%, even more preferably at least 99% identity.

25 In an even more preferred embodiment the fusion gene is a ZNF142-PTK6 fusion gene. Preferably the ZNF142-PTK6 fusion gene comprises a nucleotide sequence encoding ZNF142 (SEQ ID NO. 12) or a fragment thereof, or a protein having at least 70% identity thereto preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, yet 30 more preferably at least 95%, even more preferably at least 99% identity, and a nucleotide sequence encoding PTK6 (SEQ ID NO. 16) or a fragment thereof, or a protein having at least 70% identity thereto preferably at least 80%, more preferably at least 85%, even more

preferably at least 90%, yet more preferably at least 95%, even more preferably at least 99% identity. In a further preferred embodiment the ZNF142-PTK6 fusion gene comprises a nucleotide sequence encoding fragment of ZNF142 (SEQ ID NO. 12) or a protein having at least 70% identity thereto, preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably 99%; and nucleotide sequence encoding a fragment of PTK6 (SEQ ID NO. 16) or a protein having at least 70% identity thereto, preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99%. Preferred fragments are described herein above. In a preferred embodiment the fusion gene encodes the fusion protein comprising or consisting of ZNF142-PTK6 (SEQ ID NO. 20) or a protein having at least 70% identity thereto, a sequence having at least 70% identity thereto, preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99%. A preferred fusion gene comprises the nucleotide sequence of ZNF142-PTK6 (SEQ ID NO. 19) or a nucleotide sequence having at least 70% identity thereto, in a preferred embodiment the fusion gene consists of the nucleotide sequence of ZNF142-PTK6 (SEQ ID NO. 19), or at least 70% identity thereto preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, yet more preferably at least 95%, even more preferably at least 99% identity. In a particular preferred embodiment the ZNF142-PTK 6 fusion protein comprises at least the kinase domain having SEQ ID NO. 38, or a protein sequence having at least 70% identity thereto preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, yet more preferably at least 95%, even more preferably at least 99% identity.

The inventors found that the fusion genes according to the present invention are predictive markers for different types of lung cancer. It will be apparent that the lung cancer according to the present invention may be any lung cancer tumor types as classified in the WHO classification of lung tumors (2004), preferably selected from the group consisting of non-small cell lung cancer (NSCLC), including squamous (SQ) lung cancer, adenocarcinoma (AD) lung cancer, large cell neuroendocrine (LCNEC) lung cancer, and carcinoid (CA) lung cancer; or small cell lung cancer (SCLC). Hence, in a preferred embodiment of the method for diagnosing lung cancer the lung cancer is selected from the group consisting of non-small cell lung cancer (NSCLC), including squamous (SQ) lung cancer, adenocarcinoma (AD) lung cancer, large cell

neuroendocrine (LCNEC) lung cancer, and carcinoid (CA) lung cancer; or small cell lung cancer (SCLC). In a preferred embodiment of the invention the presence and/or expression of a SOS1-ADCY3 fusion gene is attributed to the presence of non-small cell lung cancer (NSCLC). In a preferred embodiment the presence and/or expression of a SOS1-ADCY3 and/or FNDC3B-TNIK is attributed to the presence of an adenocarcinoma. In a further embodiment the present invention the presence and/or expression of a ZNF142-PTK6 fusion gene is attributed to the presence of a lung cancer selected from the group consisting of non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC); preferably the presence and/or expression of a ZNF142-PTK6 fusion gene according to the present invention is attributed to the presence of a large-cell neuroendocrine tumor. In yet a further embodiment the presence and/or expression of an AGAP1-IGFBP2 fusion gene is attributed to the presence of large cell neuroendocrine lung cancer (LCNEC). In yet a further embodiment the presence and/or expression of a POMC-ADD1 fusion gene is attributed to the presence of lung carcinoid (CA). In a further embodiment the presence and/or expression of an FNDC3B-TNIK fusion gene is attributed to the presence of lung AD. In a further embodiment the presence and/or expression of a C12orf11-RASSF8 fusion gene is attributed to the presence of a lung AD. In a further embodiment the presence and/or expression of an E2F4-RPL14 fusion gene is attributed to the presence of a lung cancer selected from the group consisting of NSCLC and SQ lung cancer, preferably selected from the group of lung cancer consisting of FGFR1-amplified NSCLC and FGFR1-amplified SQ lung cancer. In a further embodiment the presence and/or expression of a CHEK2-THOC5 fusion gene is attributed to the presence of lung AD. In a further embodiment the presence and/or expression of an M6PR-CD4 fusion gene is attributed to the presence of LCNEC. In a further embodiment the presence and/or expression of a ZBP1-BCAS1 fusion gene is attributed to the presence of a lung AD. In a further embodiment the presence and/or expression of an EBF1-ATP10B fusion gene is attributed to the presence of LCNEC. In a further embodiment the presence and/or expression of a PIGF-CHMP3 fusion gene is attributed to the presence of lung AD. In a further embodiment the presence and/or expression of a GRLF1-DLGAP1 fusion gene is attributed to the presence of SQ lung cancer, preferably FGFR1-amplified SQ lung cancer. In a further embodiment the presence and/or expression of an E2F3-FLJ22536 fusion gene is attributed to the presence of a SQ lung cancer, preferably FGFR1-amplified SQ lung cancer. In a further embodiment the presence and/or expression of a GPR160-NCEH1 fusion gene is attributed to the presence of a SCLC.

The detection of the fusion gene may be conducted by any method as long as the presence and/or expression is detected selectively. Selective methods are known by those skilled in the art. For example the presence and expression may be detected using antibody based assays (immunoassays). However, in a preferred embodiment the detection of one or more fusion genes according to the present invention is preformed using nucleic acid detection methods. Specific nucleic acid detection methods are known by those of ordinary skills and include hybridisation and/or amplification based methods. Hence, in one embodiment of the methods according to the present invention presence and/or expression of a fusion gene according to the present invention is detected by a hybridisation and/or an amplification assay.

Other amplification methods may likewise be applied, these are for example, rolling circle amplification (such as in Liu, et al., "Rolling circle DNA synthesis: Small circular oligonucleotides as efficient templates for DNA polymerases," J. Am. Chem. Soc. 118:1587-1594 (1996).), isothermal amplification (such as in Walker, et al., "Strand displacement amplification--an isothermal, in vitro DNA amplification technique," Nucleic Acids Res. 20(7):1691-6 (1992)), ligase chain reaction (such as in Landegren, et al., "A Ligase-Mediated Gene Detection Technique," Science 241:1077-1080, 1988, or, in Wiedmann, et al., "Ligase Chain Reaction (LCR)--Overview and Applications," PCR Methods and Applications (Cold Spring Harbor Laboratory Press, Cold Spring Harbor Laboratory, NY, 1994) pp. S51-S64.)). Polymerase chain reaction amplification is preferred.

Nucleic-acid amplification can be accomplished by any of the various nucleic-acid amplification methods known in the art, including but not limited to the polymerase chain reaction (PCR), ligase chain reaction (LCR), transcription-based amplification system (TAS), nucleic acid sequence based amplification (NASBA), rolling circle amplification (RCA), transcription-mediated amplification (TMA), self-sustaining sequence replication (3SR) and Q β amplification.

Preferred methods use hybridisation and/or polymerase chain reaction (PCR) based methods. In a preferred embodiment hence the presence and/or expression of a fusion gene according to the present invention is performed using hybridisation and/or amplification assay, preferably the

hybridisation and/or amplification assay is selected from the group consisting of polymerase chain reaction (PCR), real-time PCR, reverse transcriptase PCR (RT-PCR), fluorescent *in situ* hybridisation (FISH), chromogenic *in situ* hybridisation (CISH), break-apart *in situ* hybridisation (ba-FISH), and silver *in situ* hybridisation (SISH). These methods are known by the person skilled in the art e.g. as driveable from Nitta et al., 2010; Jiang et al., 1995.

One gist of the present invention lies in the fact that a method is provided that allows for the determination of the susceptibility of a given cell, tissue or cells in a tissue, (or a cell culture or individual cells in such a cell culture, or as will be explained below, (a) cell(s) in a biological/medical/pathological sample) for anti-cancer or anti-proliferative treatment using known inhibitors.

As outlined below in the Example section herein below, it has been found that the presence or expression of a fusion gene comprising a fragment of an adenylyl cyclase results is indicative for the susceptibility of the cells to a MEK inhibitor, i.e. the cells are sensitive to MEK inhibitors. Hence, the present invention also relates to a method of selecting a cell, tissue or cell culture with susceptibility to a MEK inhibitor, comprising the steps: - determining the presence and/or expression of a fusion gene in said cell, tissue or cell culture, wherein the fusion gene comprises a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase in said cell, tissue or cell culture; and - attributing the presence and/or expression of said fusion gene to the susceptibility of the cell, tissue or cell culture to a MEK inhibitor.

Accordingly also a method to distinguish between responder and non-responder to MEK inhibitors is provided. Hence, the present invention also relates to a method for identifying a responder or non-responder to a MEK inhibitor, the method comprising the steps: - providing a sample from a patient suspected to suffer or being prone to suffer from lung cancer; and - determining the presence and/or expression of a fusion gene comprising a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase, or a fragment thereof; wherein the presence and/or expression of said fusion gene is indicative for a responder to a MEK inhibitor.

Moreover, the inventors have shown that an increased expression of a fusion gene comprising a fragment of an adenylyl cyclase triggers resistance of the cell to ALK inhibitors. Hence, the present invention also relates to a method of determining resistance of a cell, tissue or cell culture to an ALK inhibitor, comprising the steps: - determining the presence and/or expression level of a fusion gene in said cell, tissue or cell culture, wherein the fusion gene comprises a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase in said cell, tissue or cell culture; - comparing the determined expression level to the expression level in a control cell, tissue or cell culture expressing said fusion gene and not being resistant to an ALK inhibitor; and - attributing a determined expression level higher than in the control cell, tissue or cell culture to the resistance of the cell, tissue or cell culture to a ALK inhibitor; and/or - attributing a determined expression equal or lower than in the control cell, tissue or cell culture to the susceptibility of the cell, tissue or cell culture to a ALK inhibitor.

Accordingly the invention also provides for a method to distinguish between responders and non-responders to ALK inhibitors. Therefore, the present invention also relates to a method for identifying a responder or non-responder to an ALK inhibitor, the method comprising the steps: - determining the presence/expression level of a fusion gene in a sample from a patient suspected to suffer or being prone to suffer from lung cancer, said fusion gene comprising a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase, or a fragment thereof; and - comparing the determined expression level to the expression level in a sample from a control patient, wherein the control patient is a responder to ALK inhibitors and expresses said fusion gene; wherein a determined expression higher than in said sample of a control patient is indicative for a non-responder to a ALK inhibitor; and wherein a determined expression equal or lower than in said sample of a control patient is indicative for a responder to a ALK inhibitor.

ALK inhibitors are known by those skilled in the art. It will be readily understood that ALK inhibitors include all substances that inhibit the enzymatic and regulatory activities of ALK. The skilled person knows how to determine whether a substance does inhibit ALK using *in vivo* or *in vitro* techniques. The inhibition may occur differently. The activation of ALK may be inhibited. Furthermore, the activity of ALK may be inhibited by occupying the active site of the

enzyme and thereby preventing the activity (e.g. see Koivunen et al., 2008; Galkin et al., 2007). Preferred ALK inhibitors are selected from the group of ALK inhibitors consisting of TAE684, CH5424802, and Crizotinib.

5 The levels of the markers as obtained by the methods or the use of the methods according to the present invention may be analyzed in a number of fashions well known to a person skilled in the art. For example, each assay result obtained may be compared to a "normal" value, such as the sample from said control patient, or a value indicating a particular disease or outcome. A particular diagnosis/prognosis may depend upon the comparison of each assay result to such a
10 value, which may be referred to as a diagnostic or prognostic "threshold". In certain embodiments, assays for one or more diagnostic or prognostic indicators are correlated to a condition or disease by merely the presence or absence of the indicator(s) in the assay. For example, an assay can be designed so that a positive signal only occurs above a particular threshold concentration of interest, and below which concentration the assay provides no signal
15 above background.

The sensitivity and specificity of a diagnostic and/or prognostic test depends on more than just the analytical "quality" of the test, they also depend on the definition of what constitutes an abnormal result. In practice, Receiver Operating Characteristic curves (ROC curves), are
20 typically calculated by plotting the value of a variable versus its relative frequency in "normal" (i.e. apparently healthy) and "disease" populations. For any particular marker, a distribution of marker levels for subjects with and without a disease will likely overlap. Under such conditions, a test does not absolutely distinguish normal from disease with 100% accuracy, and the area of overlap indicates where the test cannot distinguish normal from disease. A threshold
25 is selected, above which (or below which, depending on how a marker changes with the disease) the test is considered to be abnormal and below which the test is considered to be normal. The area under the ROC curve is a measure of the probability that the perceived measurement will allow correct identification of a condition. ROC curves can be used even when test results don't necessarily give an accurate number. As long as one can rank results,
30 one can create a ROC curve. For example, results of a test on "disease" samples might be ranked according to degree (e.g. 1=low, 2=normal, and 3=high). This ranking can be correlated to results in the "normal" population, and a ROC curve created. These methods are well known

in the art. See, e.g., Hanley et al. 1982. Radiology 143: 29-36. Preferably, a threshold is selected to provide a ROC curve area of greater than about 0.5, more preferably greater than about 0.7, still more preferably greater than about 0.8, even more preferably greater than about 0.85, and most preferably greater than about 0.9. The term "about" in this context refers to +/- 5% of a given measurement.

The horizontal axis of the ROC curve represents (1-specificity), which increases with the rate of false positives. The vertical axis of the curve represents sensitivity, which increases with the rate of true positives. Thus, for a particular cut-off selected, the value of (1-specificity) may be determined, and a corresponding sensitivity may be obtained. The area under the ROC curve is a measure of the probability that the measured marker level will allow correct identification of a disease or condition. Thus, the area under the ROC curve can be used to determine the effectiveness of the test.

In certain embodiments, particular thresholds for one or more markers in a panel are not relied upon to determine if a profile of marker levels obtained from a subject are indicative of a particular diagnosis/prognosis. Rather, the present invention may utilize an evaluation of a marker panel "profile" as a unitary whole. A particular "fingerprint" pattern of changes in such a panel of markers may, in effect, act as a specific diagnostic or prognostic indicator. As discussed herein, that pattern of changes may be obtained from a single sample, or from temporal changes in one or more members of the panel (or a panel response value). A panel herein refers to a set of markers.

As described herein after, a panel response value is preferably determined by plotting ROC curves for the sensitivity (i.e. true positives) of a particular panel of markers versus 1-(specificity) (i.e. false positives) for the panel at various cut-offs. In these methods, a profile of marker measurements from a subject is considered together to provide a global probability (expressed either as a numeric score or as a percentage risk) of a diagnosis or prognosis. In such embodiments, an increase in a certain subset of markers may be sufficient to indicate a particular diagnosis/prognosis in one patient, while an increase in a different subset of markers may be sufficient to indicate the same or a different diagnosis/prognosis in another patient. Weighting factors may also be applied to one or more markers in a panel, for example, when a

marker is of particularly high utility in identifying a particular diagnosis/prognosis, it may be weighted so that at a given level it alone is sufficient to signal a positive result. Likewise, a weighting factor may provide that no given level of a particular marker is sufficient to signal a positive result, but only signals a result when another marker also contributes to the analysis.

5

In certain embodiments, markers and/or marker panels are selected to exhibit at least about 70% sensitivity, more preferably at least about 80% sensitivity, even more preferably at least about 85% sensitivity, still more preferably at least about 90% sensitivity, and most preferably at least about 95% sensitivity, combined with at least about 70% specificity, more preferably at least about 80% specificity, even more preferably at least about 85% specificity, still more preferably at least about 90% specificity, and most preferably at least about 95% specificity. In particularly preferred embodiments, both the sensitivity and specificity are at least about 75%, more preferably at least about 80%, even more preferably at least about 85%, still more preferably at least about 90%, and most preferably at least about 95%. The term "about" in this context refers to +/- 5% of a given measurement.

15

In other embodiments, a positive likelihood ratio, negative likelihood ratio, odds ratio, or hazard ratio is used as a measure of a test's ability to predict risk or diagnose a disease. In the case of a positive likelihood ratio, a value of 1 indicates that a positive result is equally likely among subjects in both the "diseased" and "control" groups; a value greater than 1 indicates that a positive result is more likely in the diseased group; and a value less than 1 indicates that a positive result is more likely in the control group. In the case of a negative likelihood ratio, a value of 1 indicates that a negative result is equally likely among subjects in both the "diseased" and "control" groups; a value greater than 1 indicates that a negative result is more likely in the test group; and a value less than 1 indicates that a negative result is more likely in the control group. In certain preferred embodiments, markers and/or marker panels are preferably selected to exhibit a positive or negative likelihood ratio of at least about 1.5 or more or about 0.67 or less, more preferably at least about 2 or more or about 0.5 or less, still more preferably at least about 5 or more or about 0.2 or less, even more preferably at least about 10 or more or about 0.1 or less, and most preferably at least about 20 or more or about 0.05 or less. The term "about" in this context refers to +/- 5% of a given measurement.

20

25

30

In the case of an odds ratio, a value of 1 indicates that a positive result is equally likely among subjects in both the "diseased" and "control" groups; a value greater than 1 indicates that a positive result is more likely in the diseased group; and a value less than 1 indicates that a positive result is more likely in the control group. In certain preferred embodiments, markers and/or marker panels are preferably selected to exhibit an odds ratio of at least about 2 or more or about 0.5 or less, more preferably at least about 3 or more or about 0.33 or less, still more preferably at least about 4 or more or about 0.25 or less, even more preferably at least about 5 or more or about 0.2 or less, and most preferably at least about 10 or more or about 0.1 or less. The term "about" in this context refers to $\pm 5\%$ of a given measurement.

In the case of a hazard ratio, a value of 1 indicates that the relative risk of an endpoint (e.g., death) is equal in both the "diseased" and "control" groups; a value greater than 1 indicates that the risk is greater in the diseased group; and a value less than 1 indicates that the risk is greater in the control group. In certain preferred embodiments, markers and/or marker panels are preferably selected to exhibit a hazard ratio of at least about 1.1 or more or about 0.91 or less, more preferably at least about 1.25 or more or about 0.8 or less, still more preferably at least about 1.5 or more or about 0.67 or less, even more preferably at least about 2 or more or about 0.5 or less, and most preferably at least about 2.5 or more or about 0.4 or less. The term "about" in this context refers to $\pm 5\%$ of a given measurement.

The skilled artisan will understand that associating a diagnostic or prognostic indicator, with a diagnosis or with a prognostic risk of a future clinical outcome is a statistical analysis. For example, a marker level of greater than X may signal that a patient is more likely to suffer from an adverse outcome than patients with a level less than or equal to X, as determined by a level of statistical significance. Additionally, a change in marker concentration from baseline levels may be reflective of patient prognosis, and the degree of change in marker level may be related to the severity of adverse events. Statistical significance is often determined by comparing two or more populations, and determining a confidence interval and/or a p value. See, e.g., Dowdy and Wearden, *Statistics for Research*, John Wiley & Sons, New York, 1983. Preferred confidence intervals of the invention are 90%, 95%, 97.5%, 98%, 99%, 99.5%, 99.9% and 99.99%, while preferred p values are 0.1, 0.05, 0.025, 0.02, 0.01, 0.005, 0.001, and 0.0001.

Cells carrying a fusion protein comprising a fragment of an adenylyl cyclase gene showed an increased phosphorylation status of CREB, a direct target of pPKA. Histone Acetyl Transferase (HAT) inhibitors are known inhibitors of the CREB pathway: Hence, the presence and/or expression of a fusion gene comprising a fragment of an adenylyl cyclase gene is indicative for the susceptibility of a cell to HAT inhibitors. The present invention hence also provides for a method of selecting a cell, tissue or cell culture with susceptibility to a HAT inhibitor, comprising the steps: - determining the presence and/or expression of a fusion gene in said cell, tissue or cell culture, wherein the fusion gene comprises a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase in said cell, tissue or cell culture; and - attributing the presence and/or expression of said fusion gene to the susceptibility of the cell, tissue or cell culture to a HAT inhibitor.

Accordingly, the invention also provides for a method to distinguish between responder and non-responder to a HAT inhibitor. Hence, the invention also relates to a method for identifying a responder or non-responder to a HAT inhibitor, the method comprising the steps: - determining the presence and/or expression of a fusion gene in a sample from a patient suspected to suffer or being prone to suffer from lung cancer, said fusion gene comprising a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase, or a fragment thereof; wherein the presence and/or expression of said fusion gene is indicative for a responder to a HAT inhibitor.

HAT inhibitors are known by those skilled in the art. It will be readily understood that HAT inhibitors include all substances that inhibit the enzymatic and regulatory activities of HAT. The skilled person knows how to determine whether a substance does inhibit HAT using *in vivo* or *in vitro* techniques. The inhibition may occur differently. The activation of HAT may be inhibited. Furthermore, the activity of HAT may be inhibited by occupying the active site of the enzyme and thereby preventing activity (e.g. see Bowers et al., 2010). Preferred HAT inhibitors are selected from the group of HAT inhibitors consisting of C646, and curcumin .

The inventors found that the activity of PKA is directly linked to the presence and/or expression of a fusion gene comprising a fragment of an adenylyl cyclase gene. Hence, the invention also relates to a method of selecting a cell, tissue or cell culture with susceptibility to

a PKA inhibitor, comprising the steps: - determining the presence and/or expression of a fusion gene in said cell, tissue or cell culture, wherein the fusion gene comprises a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase in said cell, tissue or cell culture; and - attributing the presence and/or expression of said fusion gene to the susceptibility of the cell, tissue or cell culture to a PKA inhibitor.

Accordingly, a method for distinguishing between responder and non-responder to a PKA inhibitor is also provided. The present invention relates to a method for identifying a responder or non-responder to a PKA inhibitor, the method comprising the steps: - determining the presence and/or expression of a fusion gene in a sample from a patient suspected to suffer or being prone to suffer from lung cancer, said fusion gene comprising a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase, or a fragment thereof; wherein the presence and/or expression of said fusion gene is indicative for a responder to a PKA inhibitor.

PKA inhibitors are known by those skilled in the art. It will be readily understood that PKA inhibitors include all substances that inhibit the enzymatic and regulatory activities of PKA. The skilled person knows how to determine whether a substance does inhibit PKA using *in vivo* or *in vitro* techniques. The inhibition may occur differently. The activation of PKA phosphorylation may be inhibited or the level of cyclic AMP (cAMP) is lowered. Furthermore, the activity of PKA may be inhibited by occupying the active site of the enzyme and thereby preventing the activity. PKA inhibitors are known by those of ordinary skills.

The determination of the presence and/or expression of a fusion gene comprising a fragment of an adenylyl cyclase gene allows the determination of susceptibility of cells to certain inhibitors. These methods may also be combined. If for example the presence and/or expression of a fusion gene comprising a nucleotide sequence encoding an adenylyl cyclase is detected an attributed to the susceptibility of the cell, tissue or cell culture to a MEK inhibitor the expression level may be detected in order to access whether the cell, tissue or cell culture is susceptible to a PKA inhibitor.

The skilled artisan will readily appreciate that the present invention therefore provides for tools

to access whether a patient suffering from lung cancer was unsuccessfully treated with one inhibitor would be a responder to another inhibitor. This allows preventing a patient from further insufficient treatments. Hence, the above outlined methods for identifying a responder or non-responder to an MEK or PKA or HAT or ALK inhibitor can be combined.

5

In one embodiment of the method for identifying a responder or non-responder to an ALK inhibitor the patient is a non-responder to PKA inhibitors and/or MEK inhibitors and/or HAT inhibitors.

10 In one embodiment of the method for identifying a responder or non-responder to a PKA inhibitor the patient is a non-responder to MEK inhibitors and/or ALK inhibitors and/or HAT inhibitors.

In one embodiment of the method for identifying a responder or non-responder to a MEK
15 inhibitor the patient is a non-responder to PKA inhibitors and/or ALK inhibitors and/or HAT inhibitors.

In one embodiment of the method for identifying a responder or non-responder to a HAT
20 inhibitor the patient is a non-responder to PKA inhibitors and/or ALK inhibitors and/or MEK inhibitors.

Therefore, the present invention does not only provide for a method for selecting cells/tissues/cell cultures which are susceptible to PKA, MEK or HAT inhibitors, but also for an *in vitro* method for assessing an individual, i.e. a human or animal patient, for its potential
25 responsiveness to an anti-cancer or anti-proliferate treatment with such inhibitors, i.e. identifying responders. The present invention provides not only for the possibility to select cells, tissues and cell cultures that are susceptible for PKA, MEK or HAT inhibitor treatment (i.e. the selection of e.g. research tools wherein drugs with a structural similarity to PKA, MEK or HAT inhibitors may be tested or which are useful in screening methods for compounds that
30 are suspected to function like known PKA, MEK or HAT inhibitors) but also for a method to evaluate whether a given patient, preferably a human patient, in need of treatment but also prevention of a proliferative disease, is a responder for inhibitor treatment. Most preferably, the

responsiveness of a given patient to PKA, MEK or HAT inhibitors is tested.

In a preferred embodiment of the present invention the adenylyl cyclase is selected from the group consisting of ADCY1 (SEQ ID NO. 141), ADCY2 (SEQ ID NO. 142), ADCY3 (SEQ ID NO. 6), ADCY4 (SEQ ID NO. 143), ADCY5 (SEQ ID NO. 144), ADCY6 (SEQ ID NO. 145), ADCY7 (SEQ ID NO. 146), ADCY8 (SEQ ID NO. 147), ADCY9 (SEQ ID NO. 148), and ADCY10 (SEQ ID NO. 149); or a protein having at least 70% identity thereto, preferably at least 80%, more preferably at least 90%, more preferably at least 95%, even more preferably at least 99% identical.

In a preferred embodiment of the methods to determine the susceptibility of a cell, cell tissue or cell culture to a MEK inhibitor or ALK inhibitor or PKA inhibitor or HAT inhibitor the fusion gene further comprises a nucleotide sequence encoding a second protein or a fragment of the second protein, said second protein being a guanine nucleotide exchange factor for RAS proteins, preferably said guanine nucleotide exchange factor for RAS proteins is SOS1 (SEQ ID NO. 2) or a protein having at least 70% identity thereto. In a preferred embodiment of these methods the fusion gene comprises one or more nucleotide sequence of a gene or a fragment of the gene, said gene having at least 70% identity, preferably at least 80%, more preferably at least 90%, more preferably at least 95%, even more preferably at least 99% identical, to a sequence selected from the group consisting of SOS1 (SEQ ID NO. 1), and ADCY3 (SEQ ID NO. 5), or a fragment thereof. Preferably the fusion gene comprises one or more nucleotide sequence of a fragment of the gene, said fragment of the gene being selected from the group consisting of SOS1 (SEQ ID NO. 3), and ADCY3 (SEQ ID NO. 7). It is further preferred that the fusion gene is a SOS1-ADCY3 fusion gene, preferably the fusion gene has a sequence at least 70% identical, preferably at least 80%, more preferably at least 90%, more preferably at least 95%, even more preferably at least 99% identical, to SOS1-ADCY3 (SEQ ID NO. 9).

As used herein, the term “cell, tissue and cell culture” is not only limited to isolated cells, tissues and cell cultures but also comprises the use of samples, i.e. biological, medical or pathological samples that consist of fluids that comprise such cells, tissues or cell cultures. Such a fluid may be a body fluid or also excrements and may also be a culture sample, like the culture medium from cultured cells or cultured tissues. The body fluids may comprise, but are

not limited to blood, serum, plasma, urine, saliva, synovial fluid, spinal fluid, cerebrospinal fluid, tears, stool and the like.

As pointed out in the technical problem above, there is a need in the art for markers, which can predict the outcome of an anti-cancer therapy with PKA, MEK and/or HAT inhibitors prior to and during treatment. There is a need for stratification of patients who are to be subjected to or are being subjected to an anti-cancer therapy with one or more of the inhibitors mentioned herein and distinguishing between “responder” and “non-responder” patients.

It has further been found that the presence and/or expression of a fusion gene comprising at least a fragment of a gene encoding for a member of the E2F family of transcription factors and/or a fragment of a gene encoding a ribosomal protein of the L14E family is indicative for the susceptibility of a tumor cell to FGFR1 inhibitors. Hence, the present invention also relates to a method of selecting a cell, tissue or cell culture with susceptibility to an FGFR1 inhibitor, comprising the steps: - determining the presence and/or expression of a fusion gene in said cell, tissue or cell culture; wherein said fusion gene comprises one or more nucleotide sequence encoding a protein or a fragment of the protein, said protein being selected from the group consisting of a member of the E2F family of transcription factors, a ribosomal protein of the L14E family; and - attributing the presence and/or expression of said fusion gene to susceptibility of the cell, tissue or cell culture to a FGFR1 inhibitor.

Furthermore, this finding allows distinguishing between responder and non-responder patients to an FGFR1 inhibitor. Hence, the present invention also relates to a method for identifying a responder or non-responder to an FGFR1 inhibitor, the method comprising the steps: - determining the presence and/or expression of a fusion gene in a sample from a patient suspected to suffer or being prone to suffer from cancer, said fusion gene comprising a nucleotide sequence encoding a protein or a fragment of the protein, said protein selected from the group consisting of a member of the E2F family of transcription factors, and a ribosomal protein of the L14E family; wherein the presence and/or expression of said fusion gene is indicative for a responder to a FGFR1 inhibitor; and/or wherein the absence and/or non expression of said fusion gene is indicative for a non-responder to a FGFR1 inhibitor.

In a preferred embodiment of the method of selecting a cell, tissue or cell culture with susceptibility to or identifying a responder or non-responder to an FGFR1 inhibitor said member of the E2F family of transcription factors is E2F4 (SEQ ID NO. 61); or a protein having at least 70% identity thereto; wherein said ribosomal protein of the L14E family is RPL14 (SEQ ID NO. 65); or a protein having at least 70% identity thereto. Preferably said fusion gene comprises one or more nucleotide sequence of a gene or a fragment of the gene, said gene having a sequence with at least 70% identity to a sequence selected from the group consisting of E2F4 (SEQ ID NO. 60), and RPL14 (SEQ ID NO. 64). In a further embodiment said fusion gene comprises one or more nucleotide sequence of a fragment of the gene, said fragment having a sequence with at least 70% identity to a sequence selected from the group consisting of E2F4 (SEQ ID NO. 62), and RPL14 (SEQ ID NO. 66). In a preferred embodiment the fusion gene comprises at least a fragment of both and hence the fusion gene is an E2F4-RPL14 fusion gene, preferably the fusion gene has a nucleotide sequence having at least 70% identity to the sequence of E2F4-RPL14 (SEQ ID NO. 68).

Furthermore, the findings of the present invention provide for a tool to test tumor cells for their susceptibility to inhibitors of protein tyrosine kinase inhibitors. Hence, the invention also relates to a method of selecting a cell, tissue or cell culture with susceptibility to a PTK inhibitor, comprising the steps: - determining the presence and/or expression of a fusion gene in said cell, tissue or cell culture, wherein the fusion gene comprises a nucleotide sequence encoding a protein or a fragment of the protein, said protein being a protein tyrosine kinase, in said cell, tissue or cell culture; and - attributing the presence and/or expression of said fusion gene to the susceptibility of the cell, tissue or cell culture to a PTK inhibitor.

Accordingly the invention also provides for a tool to distinguish between patients that are responder or non-responder to a PTK inhibitor. Hence, the invention relates to a method for identifying a responder or non-responder to a PTK inhibitor, the method comprising the steps: - determining the presence and/or expression of a fusion gene in a sample from a patient suspected to suffer or being prone to suffer from lung cancer, said fusion gene comprising a nucleotide sequence encoding a protein or a fragment of the protein, said protein being a protein tyrosine kinase, or a fragment thereof; wherein the presence and/or expression of said fusion gene is indicative for a responder to a PTK inhibitor. In a preferred embodiment said

protein tyrosine kinase is PTK6 (SEQ ID NO. 16) or a protein having at least 70% identity thereto, preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably 99%; and the PTK inhibitor is a PTK6 inhibitor.

- 5 In a further preferred embodiment the fusion gene beside the above mentioned PTK, e.g. PTK6, further comprises nucleotide sequence encoding a protein or a fragment of the protein, said protein being a Zinc finger protein, preferably the Zinc finger protein is ZNF142 (SEQ ID NO. 12) or a protein having at least 70% identity thereto, preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99%. An especially preferred fusion gene encodes a protein sequence of ZNF142-PTK6 (SEQ
10 ID NO. 20) or a protein having at least 70% identity thereto, preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99%.
- 15 The inventors further found that the presence of a fusion gene comprising at least a fragment of a germinal center kinase gene in a cell correlates with the susceptibility of a tumor cell to c-Jun inhibitors. Hence, the invention also encompasses a method of selecting a cell, tissue or cell culture with susceptibility to a c-Jun inhibitor, comprising the steps: - determining the presence and/or expression of a fusion gene in said cell, tissue or cell culture, wherein the fusion gene
20 comprises a nucleotide sequence encoding a protein or a fragment of the protein, said protein being a germinal center kinase, in said cell, tissue or cell culture; and - attributing the presence and/or expression of said fusion gene to the susceptibility of the cell, tissue or cell culture to a c-Jun inhibitor.
- 25 Accordingly the present invention also relates to a method for identifying a responder or non-responder to a c-Jun inhibitor, the method comprising the steps: - determining the presence and/or expression of a fusion gene in a sample from a patient suspected to suffer or being prone to suffer from lung cancer, said fusion gene comprising a nucleotide sequence encoding a protein or a fragment of the protein, said protein being germinal center kinase, or a fragment
30 thereof; wherein the presence and/or expression of said fusion gene is indicative for a responder to a c-Jun inhibitor. In one embodiment the germinal center kinase is TNIK (SEQ ID NO. 45) or a protein having at least 70% identity thereto, preferably at least 80% identity,

more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99%. In a preferred embodiment the fusion gene in addition to a germinal center kinase gene or fragment further comprises nucleotide sequence encoding a protein or a fragment of the protein, said protein being fibronectin type III domain containing protein, preferably said
5 fibronectin type III domain containing protein is FNDC3B (SEQ ID NO. 41) or a protein having at least 70% identity thereto. Preferably said fusion gene encodes a protein sequence of FNDC3B-TNIK (SEQ ID NO 49) or a protein having at least 70% identity thereto, preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99%.

10 Furthermore, there is a need to streamline anti cancer therapy using these inhibitors. One gist of the present invention is the provision of tools to predict the outcome of a treatment with one or more of the inhibitors and to decide for the drug to be used. The methods for identifying a responder or non-responder to a MEK inhibitor and/or ALK inhibitor and/or HAT inhibitor
15 and/or PKA inhibitor and/or FGFR1 inhibitor and/or PTK inhibitor and/or c-Jun inhibitor may be combined. For example the presence and/or expression and/or expression level of a fusion gene comprising a nucleotide sequence encoding guanine nucleotide exchange factor for RAS proteins may be determined in order to identify a responder or non-responder to a MEK inhibitor and/or ALK inhibitor and/or HAT inhibitor and/or PKA inhibitor and/or PTK
20 inhibitor; and (in parallel) the presence and/or expression and/or expression level of a fusion gene comprising a nucleotide sequence encoding a germinal center kinase (as outlined above) may be determined in order to identify a responder or non-responder to a c-Jun inhibitor. Or the presence and/or expression and/or expression level of a fusion gene comprising a nucleotide sequence encoding guanine nucleotide exchange factor for RAS proteins may be
25 determined in order to identify a responder or non-responder to a MEK inhibitor and/or ALK inhibitor and/or HAT inhibitor and/or PKA inhibitor and/or PTK inhibitor and in parallel the presence and/or expression and/or expression level of a fusion gene comprising a nucleotide sequence encoding germinal center kinase may be determined in order to identify a responder or non-responder to a c-Jun inhibitor.

30 For all the methods of selecting a cell, tissue or cell culture with susceptibility to a and/or methods for identifying a responder or non-responder to a certain inhibitor using the

determination of the presence and/or expression it is one embodiment that the expression of the respective fusion gene is determined by comparing its expression level in the sample to be tested to a control sample. If as outlined above for the respective methods, the presence and/or expression of a certain fusion gene is correlated or attributed to the susceptibility to an inhibitor or to the responder for said inhibitor, this in one embodiment means that an expression level of the fusion gene higher than in a control is correlated or attributed to the susceptibility to an inhibitor or to the responder for said inhibitor. Accordingly if the absence and/or non-expression of a certain fusion gene is correlated or attributed to the susceptibility to an inhibitor or to the responder for said inhibitor, this in one embodiment means that an expression level of the fusion gene lower than in a control is correlated or attributed to the susceptibility to an inhibitor or to the responder for said inhibitor.

In context with the present invention, “different”, “higher” or “lower” means different, higher or lower than the normal (range of) expression of a fusion gene. For example, different, higher or lower means at least 1.5 fold, at least 2 fold, at least 2.5 fold, at least 3 fold, at least 4 fold, at least 5 fold, at least 7 fold, at least 10 fold, at least 15 fold, at least 25 fold, at least 50 fold, at least 100 fold, at least 200 fold different, higher or lower, wherein the higher values are preferred.

Whether the activity or expression level changes (i.e., in a certain direction like “higher” or “lower”) and to which extent the expression level of a fusion gene changes can easily be deduced by the skilled person based on the teaching provided herein and the common general knowledge. As pointed out herein, in addition to the expression level of one fusion gene also the expression level of one or further fusion genes may be measured or assessed. This optional measurement or assessment of further fusion genes’ expression levels, can also be easily deduced by the skilled person based on the teaching provided herein and the common general knowledge.

It is particularly preferred in this context that the control subject/patient is subjected to the same treatment of the cancer described and defined herein as the subject/patient itself and/or that it is known whether the control subject/patient is a responder or non-responder to this treatment.

Furthermore, the methods according to the present invention provide for a tool to monitor the response to a cancer treatment/therapy. Whether a subject/patient is a “responder” or “non-responder” with respect to a certain kind of cancer treatment/therapy can be evaluated by the skilled person on the basis of his common general knowledge and/or the teaching provided herein.

Furthermore, as the presence of a fusion gene according to the present invention is correlated with the presence of lung cancer, i.e. these fusion genes are oncogenic, the present invention also provides methods for evaluating the response of a patient to cancer treatment, comprising the step of determining the expression level of a fusion gene according to the present invention as outlined above in a sample from said patient. It is apparent to the skilled artisan that the patient responds to cancer treatment/therapy, if expression the fusion gene or its downstream targets is reduced upon said treatment/therapy. Preferably, the expression is reduced to control expression (e.g. determined in a sample obtained from a person not suffering from said cancer).

In other words, a reduction in expression level of a fusion gene or its downstream targets according to the present invention is indicative for a successful treatment/therapy. A skilled person is readily in the position to determine whether a patient responds to cancer treatment/therapy by evaluation of the expression level of one or more fusion gene or its/their downstream targets according to the present invention. In addition to the evaluation of said expression level, a person skilled in the art may also determine cytological/haematological parameters characteristic for a specific lung cancer in order to assess whether a patient responds to cancer treatment/therapy.

In contrast, a patient who does not respond to treatment/therapy does not show a reduced expression level of the fusion gene and/or its downstream targets according to the present invention upon said treatment/therapy. This is in contrast to “responders/responding patients”, showing such a reduced expression level.

In particular, a “responder” may be a subject/patient whose cytological/haematological parameters and/or (aberrant) fusion gene/downstream targets expression level change towards the “normal” gene expression level(s) or to non expression and/or loss of the fusion gene (in a sufficient manner) upon the cancer treatment/therapy. In one specific embodiment, a

“responder” may be a subject/patient not suffering from one of the herein defined resistances. In particular, a “non-responder” may be a subject/patient whose cytological/haematological parameters and/or (aberrant) expression level of one or more fusion genes/downstream targets according to the present invention do not change towards their “normal” or total loss of (expression) level(s) (in a sufficient manner) upon the cancer treatment/therapy. In one specific embodiment, a “non-responder” may be a subject/patient suffering from one of the herein defined resistances.

In context of this embodiment of the invention, one non-limiting example of a (diseased) control subject/patient (responder and/or non-responder) suffering from a cancer defined herein or being prone to suffering from a susceptibility thereto is one having and/or expressing a fusion gene comprising one or more nucleotide sequences encoding a protein or a fragment of the protein, said protein being selected from the group consisting of a guanine nucleotide exchange factor for RAS proteins, and an adenylyl cyclase, preferably comprising two nucleotide sequences encoding a protein or a fragment of the protein each, said proteins being a guanine nucleotide exchange factor for RAS proteins, and an adenylyl cyclase, preferably with the embodiments of the respective fusion gene as set out herein above.

The skilled person is aware of how a typical/desired response to a known therapy/treatment of a known lung cancer characterized by the presence and/or expression of at least one fusion gene according to the present invention (e.g. non-small cell lung cancer, lung adenocarcinoma, pancreatic cancer, colorectal cancer) should proceed or is intended to proceed. Moreover, the skilled person can consider how a typical/desired response to a (unknown) therapy/treatment of a (unknown) cancer characterized by the presence of at least one fusion gene according to the present invention proceeds or is intended to proceed. Based on this knowledge, the means, methods and uses of this invention referring to the efficacy of a therapy/treatment of such a cancer can, for example, also be carried out without employing (a sample of) a particular control subject/patient, i.e. without comparing the expression level of a fusion gene according to the present invention with a reference or control expression level determined in (a sample from) a control subject/patient. Simply by comparing the course of the determined “expression level of the fusion gene during the therapy/treatment of a cancer with the above-mentioned known “typical/desired response”, the skilled person is able to consider about the efficacy of

the therapy/treatment monitored/predicted. If the response of a subject/patient is as fast (or even faster) than the “typical/desired response”, the subject/patient is a “responder”. If the response of a subject/patient is slower than the “typical/desired response”, the subject/patient is a “non-responder” (when no substantial response can be seen) or “weak-responder”.

5 In general, a (desired) efficacy of a treatment of a cancer described herein or susceptibility thereto is indicated/predicted, when the aberrant (i.e. enhanced or decreased) expression level of the fusion gene shifted down towards the “normal level” of a (healthy) control subject/patient due to/in consequence of said treatment of the cancer or susceptibility thereto.

10 In context of this invention, the efficacy of a treatment of the cancer defined herein is high, when the subject/patient (to be) treated responds as fast (or even faster) and as complete as a “responder”, i.e. exhibits a “typical/desired response”. This means that said subject/patient reaches the “normal” level of the relevant cytological/haematological parameters and/or
15 (fusion) gene/downstream targets expression level(s) of a healthy subject/patient as fast as a “responder”, i.e. in the same manner as in a “typical/desired response”.

In context of this invention, the efficacy of a treatment of the cancer defined herein is moderate/low, when the subject/patient (to be) treated responds not as fast and/or not as
20 complete as a “responder”, i.e. does not exhibit a “typical/desired response”. This means that said subject/patient does not reach the “normal” level of the relevant cytological/haematological parameters and/or expression level of one or more fusion genes (or its/their downstream targets) according to the present invention. Accordingly, a moderate/low efficacy means also that the expression level of the fusion gene of a healthy subject/patient is
25 not reached as complete and/or as fast as a “responder”, i.e. not in the same manner as in a “typical/desired response”.

In context of this invention, there is no efficacy of a treatment of the cancer at all, when the subject/patient (to be) treated does not respond at all.

30 Particularly, when the efficacy of a treatment of the cancer as monitored/predicted in context of this invention is moderate/low or, if there is no such efficacy, a change of the (planned) therapy/treatment might be considered.

In an alternative embodiment, of the herein disclosed means, methods and uses of monitoring/predicting/identifying, the reference expression or control expression level of a fusion gene according to the present invention of a control sample/subject/patient can be replaced by a reference expression level from the subject/patient to be treated itself obtained prior to (or at the beginning of) the treatment/therapy or a sample is from healthy tissue (non-cancer tissue) of the subject/patient. In this specific case, the "control subject/patient" would be the subject/patient to be treated itself. The efficacy of the cancer treatment would then be assessed on the basis of how the expression level of the at least one fusion gene in accordance with this invention changes during treatment/therapy compared with said particular reference expression level. The more significant and/or faster said change is the more efficacious is the treatment/therapy.

It is of note that the skilled person is readily in the position to elect (a) suitable control patient(s)/subject(s)/or tissue(s) and the point(s) in time when the (reference) expression level are determined for each individual setup of the means, methods and uses provided.

Furthermore, as outlined herein, some of the identified fusion genes are activating fusion genes. As this activity is related the oncogenicity it is apparent for the skilled person consulting his common general knowledge and the teaching disclosed herein that these fusion genes are targets for anti-cancer drugs. Hence, the present invention also relates to a method for identifying compounds for the treatment of lung cancer, the method comprising the steps: - providing a cell expressing a fusion gene, wherein the fusion gene comprises one or more nucleotide sequence encoding a protein or a fragment of the protein, said protein being selected from the group consisting of a guanine nucleotide exchange factor for RAS proteins, an adenylyl cyclase, a Zink finger protein, a protein tyrosine kinase, a member of ADP-ribosylation factor GTPase activating (ARF-GAP) protein family, a member of the IGF-1R pathway, a polypeptide hormone precursors, an adducin, a fibronectin type III domain containing protein, a germinal center kinase (GCKs), a chromosome open reading frame, a member of the Ras-association domain family (RASSF), a member of the E2F family of transcription factors, a ribosomal protein of the L14E family; - contacting said cell to a candidate compound; and - evaluating the viability of said cell contacted with said candidate compound.

It will be apparent those of ordinary skills that the method of identifying compounds for the treatment of lung cancer may be performed using any suited cell line. Preferably the cell line is a lung cancer cell-line.

5

In a preferred embodiment of the invention the method of identifying compounds for the treatment of lung cancer said guanine nucleotide exchange factor for RAS proteins is SOS1 (SEQ ID NO. 2) or a protein having at least 70% identity thereto; wherein said adenylyl cyclase is selected from the group of adenylyl cyclises consisting of ADCY1 (SEQ ID NO. 141),
10 ADCY2 (SEQ ID NO. 142), ADCY3 (SEQ ID NO. 6) ADCY4 (SEQ ID NO. 143), ADCY5 (SEQ ID NO. 144), ADCY6 (SEQ ID NO. 145), ADCY7 (SEQ ID NO. 146), ADCY8 (SEQ ID NO. 147), ADCY9 (SEQ ID NO. 148), and ADCY10 (SEQ ID NO. 149); or a protein having at least 70% identity thereto; wherein said Zink finger protein is ZNF142 (SEQ ID NO. 12) or a protein having at least 70% identity thereto; wherein said protein tyrosine kinase is
15 PTK6 (SEQ ID NO. 16) or a protein having at least 70% identity thereto; wherein said member of ADP-ribosylation factor GTPase activating (ARF-GAP) protein family is AGAP1 (SEQ ID NO. 22) or a protein having at least 70% identity thereto; wherein said member of the IGF-1R pathway is IGFBP2 (SEQ ID NO. 26) or a protein having at least 70% identity thereto; wherein said polypeptide hormone precursor is POMC (SEQ ID NO. 32) or a protein having at least
20 70% identity thereto; wherein said adducin is ADD1 (SEQ ID NO. 36) or a protein having at least 70% identity thereto; wherein said fibronectin type III domain containing protein is FNDC3B (SEQ ID NO. 41) or a protein having at least 70% identity thereto; wherein said germinal center kinase (GCK) is TNIK (SEQ ID NO. 45) or a protein having at least 70% identity thereto; wherein chromosome open reading frame is C12orf11 (SEQ ID NO. 51) or a
25 protein having at least 70% identity thereto; wherein said member of the Ras-association domain family (RASSF) is RASSF8 (SEQ ID NO. 55) or a protein having at least 70% identity thereto; wherein said member of the E2F family of transcription factors is E2F4 (SEQ ID NO. 61), or a protein having at least 70% identity thereto; and wherein said ribosomal protein of the L14E family is RPL14 (SEQ ID NO. 65) or a protein having at least 70% identity thereto. In a
30 yet more preferred embodiment said fusion gene comprises one or more nucleotide sequence of a gene or a fragment of the gene, said gene having a sequence at least 70% identity to a sequence selected from the group consisting of SOS1 (SEQ ID NO. 1), ADCY3 (SEQ ID NO.

5), ZNF142 (SEQ ID NO. 11), PTK6 (SEQ ID NO. 15), AGAP1 (SEQ ID NO. 21), IGFBP2 (SEQ ID NO. 25), POMC (SEQ ID NO. 31), ADD1 (SEQ ID NO. 35), FNDC3B (SEQ ID NO. 40), TNIK (SEQ ID NO. 44), C12orf11 (SEQ ID NO. 50), RASSF8 (SEQ ID NO. 54), E2F4 (SEQ ID NO. 60), and RPL14 (SEQ ID NO. 64). As outlined herein certain fragments of the genes are preferred in context with the fusion genes according to the present invention. Hence the fusion gene in a further embodiment comprises one or more nucleotide sequence of a fragment of the gene, said fragment having a sequence at least 70% identity to a sequence selected from the group consisting of SOS1 (SEQ ID NO. 3), ADCY3 (SEQ ID NO. 7), ZNF142 (SEQ ID NO. 13), PTK6 (SEQ ID NO. 17), AGAP1 (SEQ ID NO. 23), IGFBP2 (SEQ ID NO. 27), POMC (SEQ ID NO. 33), ADD1 (SEQ ID NO. 37), FNDC3B (SEQ ID NO. 42), TNIK (SEQ ID NO. 46), C12orf11 (SEQ ID NO. 52), RASSF8 (SEQ ID NO. 56), E2F4 (SEQ ID NO. 62), and RPL14 (SEQ ID NO. 66). It is also preferred that said fusion gene is a fusion gene selected from the group of fusion genes consisting of SOS1-ADCY3, ZNF142-PTK6, AGAP1-IGFBP2, POMC-ADD1, FNDC3B-TNIK, C12orf11-RASSF8, and E2F4-RPL14, preferably the fusion gene has a sequence with at least 70% identity to a sequence selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 9), ZNF142-PTK6 (SEQ ID NO. 19), AGAP1-IGFBP2 (SEQ ID NO. 29), POMC-ADD1 (SEQ ID NO. 39), FNDC3B-TNIK (SEQ ID NO. 48), C12orf11-RASSF8 (SEQ ID NO. 58), and E2F4-RPL14 (SEQ ID NO. 68).

20 In a particular preferred embodiment of the invention the method of identifying compounds for the treatment of lung cancer said guanine nucleotide exchange factor for RAS proteins is SOS1 (SEQ ID NO. 2) or a protein having at least 70% identity thereto, preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably 99%; wherein said adenylyl cyclase is ADCY3 (SEQ ID NO. 6) or a protein having at least 70% identity thereto, preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably 99%; wherein said Zink finger protein is ZNF142 (SEQ ID NO. 12) or a protein having at least 70% identity thereto, preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably 99%; wherein said protein tyrosine kinase is PTK6 (SEQ ID NO. 16) or a protein having at least 70% identity thereto, preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably 99%; wherein said fibronectin type III domain containing protein is FNDC3B (SEQ ID NO. 41)

or a protein having at least 70% identity thereto, preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably 99%; wherein said germinal center kinase (GCK) is TNIK (SEQ ID NO. 45) or a protein having at least 70% identity thereto, preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99%. In a yet more preferred embodiment said fusion gene comprises one or more nucleotide sequence of a gene or a fragment of the gene, said gene having a sequence at least 70% identity to a sequence selected from the group consisting of SOS1 (SEQ ID NO. 1), ADCY3 (SEQ ID NO. 5), ZNF142 (SEQ ID NO. 11), PTK6 (SEQ ID NO. 15), FNDC3B (SEQ ID NO. 40), and TNIK (SEQ ID NO. 44). As outlined herein certain fragments of the genes are preferred in context with the fusion genes according to the present invention. Hence the fusion gene in a further embodiment comprises one or more nucleotide sequence of a fragment of the gene, said fragment having a sequence at least 70% identity to a sequence selected from the group consisting of SOS1 (SEQ ID NO. 3), ADCY3 (SEQ ID NO. 7), ZNF142 (SEQ ID NO. 13), PTK6 (SEQ ID NO. 17), FNDC3B (SEQ ID NO. 42), and TNIK (SEQ ID NO. 46), preferably at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99%. It is also preferred that said fusion gene is a fusion gene selected from the group of fusion genes consisting of SOS1-ADCY3, ZNF142-PTK6, and FNDC3B-TNIK, preferably the fusion gene has a sequence with at least 70% identity to a sequence selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 9), ZNF142-PTK6 (SEQ ID NO. 19), and FNDC3B-TNIK (SEQ ID NO. 48).

The present invention furthermore relates to the use of a cell line expressing at least one fusion gene according to the present invention, preferably a fusion genes selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 9), ZNF142-PTK6 (SEQ ID NO. 19), AGAP1-IGFBP2 (SEQ ID NO. 29), POMC-ADD1 (SEQ ID NO. 39), FNDC3B-TNIK (SEQ ID NO. 48), C12orf11-RASSF8 (SEQ ID NO. 58), and E2F4-RPL14 (SEQ ID NO. 68) for screening and/or validation of a medicament for the treatment of lung cancer, preferably the cell line is a transgenic cell line. In a particular preferred embodiment the invention furthermore relates to the use of a cell line expressing at least one fusion gene according to the present invention, wherein the fusion genes selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 9), ZNF142-PTK6 (SEQ ID NO. 19), and FNDC3B-TNIK (SEQ ID NO. 48), C12orf11-

RASSF8 (SEQ ID NO. 58), and E2F4-RPL14 (SEQ ID NO. 68) for screening and/or validation of a medicament for the treatment of lung cancer, preferably the cell line is a transgenic cell line.

- 5 “Lung cancer” in context with the present invention refers to any proliferative disease of the lung or lung cells, preferably selected from the group of lung cancer types consisting of squamous, adenocarcinoma, small cell carcinoma, large cell carcinoma (including large cell neuroendocrine carcinoma), and carcinoid tumor.
- 10 The invention also relates to nucleic acids comprising a sequence at least 70% identical to a sequence selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 9), ZNF142-PTK6 (SEQ ID NO. 19), AGAP1-IGFBP2 (SEQ ID NO. 29), POMC-ADD1 (SEQ ID NO. 39), FNDC3B-TNIK (SEQ ID NO. 48), C12orf11-RASSF8 (SEQ ID NO. 58), E2F4-RPL14 (SEQ ID NO. 68), CHEK2-THOC5 (SEQ ID NO. 77), M6PR-CD4 (SEQ ID NO. 86), ZBP1-BCAS1
- 15 (SEQ ID NO. 95), EBF1-ATP10B (SEQ ID NO. 104), PIGF-CHMP3 (SEQ ID NO. 113), GRLF1-DLGAP1 (SEQ ID NO. 122), E2F3-FLJ22536 (SEQ ID NO. 130), and GPR160-NCEH1 (SEQ ID NO. 139). In a particular preferred embodiment the invention relates to a nucleic acid comprising a sequence at least 70% identical to a sequence selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 9), ZNF142-PTK6 (SEQ ID NO. 19), and
- 20 FNDC3B-TNIK (SEQ ID NO. 48). As the inventors found that the ZNF142-PTK6 fusion gene encodes for a new kinase domain (SEQ ID NO. 38), The invention also relates to a nucleic acid coding for a protein, said protein comprising a sequence at least 70% identical to the sequence of SEQ ID NO. 38. It will be readily understood that the embodiments of the nucleic acids according to the invention also comprise sequences encoding a protein, said protein comprising
- 25 a sequence preferably least 80% identical to SEQ ID NO. 38, more preferably at least 90% identical, even more preferably at least 95%, yet more preferably at least 99% identical to the recited sequences.
- 30 Furthermore, also a nucleic acid is encompassed, wherein the nucleic acid is encoding a sequence at least 70% identical to a protein sequence selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 10), ZNF142-PTK6 (SEQ ID NO. 20), AGAP1-IGFBP2 (SEQ ID

NO. 30), FNDC3B-TNIK (SEQ ID NO. 49), C12orf11-RASSF8 (SEQ ID NO. 59), E2F4-RPL14 (SEQ ID NO. 69), CHEK2-THOC5 (SEQ ID NO. 78), M6PR-CD4 (SEQ ID NO. 87), ZBP1-BCAS1 (SEQ ID NO. 96), EBF1-ATP10B (SEQ ID NO. 105), , GRLF1-DLGAP1 (SEQ ID NO. 123), and E2F3-FLJ22536 (SEQ ID NO. 131). In a preferred embodiment the nucleic acid comprises a sequence having at least 70% identity to a sequence selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 10), ZNF142-PTK6 (SEQ ID NO. 20), and FNDC3B-TNIK (SEQ ID NO. 49); preferably at least 80% identity , more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99% identity. Preferably, the nucleic acid comprises a sequence having at least 70% identity to SOS1-ADCY3 (SEQ ID NO. 10), preferably at least 80% identity , more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99%. In a further preferred embodiment the nucleic acid comprises a sequence having at least 70% identity to ZNF142-PTK6 (SEQ ID NO. 20), preferably at least 80% identity , more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99% identity. In a likewise preferred embodiment the nucleic acid comprises a sequence having at least 70% identity to ZNF142-PTK6 (SEQ ID NO. 20), preferably at least 80% identity , more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99% identity.

The invention moreover relates to a protein having a sequence with at least 70 % identity to a sequence selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 10), ZNF142-PTK6 (SEQ ID NO. 20), AGAP1-IGFBP2 (SEQ ID NO. 30), FNDC3B-TNIK (SEQ ID NO. 49), C12orf11-RASSF8 (SEQ ID NO. 59), E2F4-RPL14 (SEQ ID NO. 69), CHEK2-THOC5 (SEQ ID NO. 78), M6PR-CD4 (SEQ ID NO. 87), ZBP1-BCAS1 (SEQ ID NO. 96), EBF1-ATP10B (SEQ ID NO. 105), GRLF1-DLGAP1 (SEQ ID NO. 123), and E2F3-FLJ22536 (SEQ ID NO. 131), preferably having at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99% identity. In a particular preferred embodiment the invention relates to a protein having a sequence having at least 70 % identity to a sequence selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 10), ZNF142-PTK6 (SEQ ID NO. 20), and FNDC3B-TNIK (SEQ ID NO. 49), preferably having at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99% identity. In further preferred embodiment the invention relates to a

protein having at least 70 % identity to the sequence of SOS1-ADCY3 (SEQ ID NO. 10), preferably having at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99% identity. In likewise preferred embodiment the invention relates to a protein having at least 70 % identity to the sequence of

5 ZNF142-PTK6 (SEQ ID NO. 20), preferably having at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99% identity. In likewise preferred embodiment the invention relates to a protein having at least 70 % identity to the sequence of FNDC3B-TNIK (SEQ ID NO. 49), preferably having at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more

10 preferably at least 99% identity. It is also preferred that the protein comprises a sequence at having at least 70 % identity to SEQ ID NO. 38, preferably having at least 80% identity, more preferably at least 90% identity, even more preferably at least 95%, yet more preferably at least 99% identity.

15 In one embodiment the present invention also relates to a kit useful for carrying out the method comprising a step of identifying a fusion gene according to the present invention, said kit comprising an oligonucleotide or polynucleotide capable of amplifying and/or specifically detecting a fusion gene selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 9), ZNF142-PTK6 (SEQ ID NO. 19), AGAP1-IGFBP2 (SEQ ID NO. 29), POMC-ADD1 (SEQ ID

20 NO. 39), FNDC3B-TNIK (SEQ ID NO. 48), C12orf11-RASSF8 (SEQ ID NO. 58), E2F4-RPL14 (SEQ ID NO. 68), CHEK2-THOC5 (SEQ ID NO. 77), M6PR-CD4 (SEQ ID NO. 86), ZBP1-BCAS1 (SEQ ID NO. 95), EBF1-ATP10B (SEQ ID NO. 104), PIGF-CHMP3 (SEQ ID NO. 113), GRLF1-DLGAP1 (SEQ ID NO. 122), E2F3-FLJ22536 (SEQ ID NO. 130), and GPR160-NCEH1 (SEQ ID NO. 139); and/or said kit comprising an antibody specifically

25 binding to a fusion protein selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 10), ZNF142-PTK6 (SEQ ID NO. 20), AGAP1-IGFBP2 (SEQ ID NO. 30), FNDC3B-TNIK (SEQ ID NO. 49), C12orf11-RASSF8 (SEQ ID NO. 59), E2F4-RPL14 (SEQ ID NO. 69), CHEK2-THOC5 (SEQ ID NO. 78), M6PR-CD4 (SEQ ID NO. 87), ZBP1-BCAS1 (SEQ ID NO. 96), EBF1-ATP10B (SEQ ID NO. 105), GRLF1-DLGAP1 (SEQ ID NO. 123), and E2F3-

30 FLJ22536 (SEQ ID NO. 131). Preferred kits comprise an oligonucleotide or polynucleotide capable of amplifying and/or specifically detecting SOS1-ADCY3 (SEQ ID NO. 9) and/or an antibody specifically binding to SOS1-ADCY3 (SEQ ID NO. 10).

In one embodiment the present invention also relates to a kit useful for carrying out the method comprising a step of identifying a fusion gene according to the present invention, said kit comprising an oligonucleotide or polynucleotide capable of amplifying and/or specifically
5 detecting a fusion gene selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 9), ZNF142-PTK6 (SEQ ID NO. 19), and FNDC3B-TNIK (SEQ ID NO. 48); and/or said kit comprising an antibody specifically binding to a fusion protein selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 10), ZNF142-PTK6 (SEQ ID NO. 20), SEQ ID NO. 38, and FNDC3B-TNIK (SEQ ID NO. 49).

10 The skilled artisan using his general common knowledge and the teaching of the present invention is able to decide on primers or probes specifically amplifying and/or detecting the fusion genes. For example the skilled person may use a primer specifically binding to the nucleotide sequence of the 5'-fusion partner and combine it with a primer specifically binding
15 to the 3'-fusion partner in a reverse complement manner and use these to amplify at least a fragment of the fusion gene. Furthermore the skilled artisan may choose a hybridisation probe specifically binding at the sequence at which the two fusion partners are connected, i.e. at the junction of the two fusion partner. However, the kit may also comprise two hybridisation probes which are differentially labelled, e.g. for the use in a ba-FISH assay. In such case one
20 probe specifically hybridises to the sequence of one fusion partner while the other probe hybridises to a second fusion partner. The proximity of the two signals indicates the proximity of the two partners and, if close, suggesting the presence of the translocation. The proximity is e.g. detected by fluorescence resonance energy transfer (FRET).

25 "Specific Hybridization" in context with the present invention means that the nucleic acid primer or probe hybridises to the mentioned target nucleic acid under stringent conditions. Hybridization conditions are known to those skilled in the art and can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y., 6.3.1-6.3.6, 1991. Stringent conditions are in a preferred embodiment defined as equivalent to hybridization in 6X sodium
30 chloride/sodium citrate (SSC) at 45°C, followed by a wash in 0.2 X SSC, 0.1 % SDS at 65°C.

The term "antibody" comprises monoclonal and polyclonal antibodies and binding fragments

thereof, in particular Fc-fragments as well as so called "single-chain-antibodies" (Bird R. E. et al (1988) Science 242:423-6), chimeric, humanized, in particular CDR-grafted antibodies, and dia or tetrabodies (Holliger P. et al (1993) Proc. Natl. Acad. Sci. U.S.A. 90:6444-8). Also comprised are immunoglobulin like proteins that are selected through techniques including, for example, phage display to specifically bind to the molecule of interest contained in a sample. In this context the term "specific binding" refers to antibodies raised against the molecule of interest or a fragment thereof. An antibody is considered to be specific, if its affinity towards the molecule of interest or the aforementioned fragment thereof is at least 50-fold higher, preferably 100-fold higher, more preferably at least 1000-fold higher than towards other molecules comprised in a sample containing the molecule of interest. It is well known in the art how to make antibodies and to select antibodies with a given specificity.

The inventors found that the expression of a fusion protein comprising at least a fragment of an adenylyl cyclase results in a higher level of phosphorylated PKA and/or CREB. Hence, the phosphorylation status of PKA and/or CREB is a marker for the presence and expression of such fusion genes. Hence, the present invention also relates to a method for identifying the presence and expression of a fusion gene comprising a nucleic acid encoding an adenylyl cyclase or a fragment thereof in a sample, the method comprising the step of: (i) determining the phosphorylation status of PKA and/or CREB; (ii) comparing the determined phosphorylation status to the phosphorylation status of PKA and/or CREB in a control; and (iii) attributing a higher phosphorylation status of PKA and/or CREB in determined step (i) than in the control to the presence and expression of a fusion gene comprising a nucleotide sequence encoding an adenylyl cyclase or a fragment thereof. Preferred adenylyl cyclases are selected from the group consisting of ADCY1 (SEQ ID NO. 141), ADCY2 (SEQ ID NO. 142), ADCY3 (SEQ ID NO. 6) ADCY4 (SEQ ID NO. 143), ADCY5 (SEQ ID NO. 144), ADCY6 (SEQ ID NO. 145), ADCY7 (SEQ ID NO. 146), ADCY8 (SEQ ID NO. 147), ADCY9 (SEQ ID NO. 148), and ADCY10 (SEQ ID NO. 149); or a protein having at least 70% identity thereto. In a more preferred embodiment the adenylyl cyclase is ADCY3 (SEQ ID NO. 6) or a protein having at least 70% identity thereto. In a further preferred embodiment a higher phosphorylation status of PKA and/or CREB determined in step (i) than in the control is attributed to the presence and expression of a fusion gene encoding the sequence of a fragment of ADCY3 (SEQ ID NO. 6) or a protein having at least 70% identity thereto.

In a further embodiment a higher phosphorylation status of PKA and/or CREB determined in step (i) than in the control is attributed to the presence and expression of a fusion gene further comprising a nucleotide sequence encoding a second protein or a fragment of the second protein, said second protein being a guanine nucleotide exchange factor for RAS proteins. Preferably the protein being a guanine nucleotide exchange factor for RAS proteins is SOS1 (SEQ ID NO. 2), or a protein having at least 70% identity thereto.

In a preferred embodiment a higher phosphorylation status of PKA and/or CREB determined in step (i) than in the control is attributed to the presence and expression of a fusion gene is a SOS1-ADCY3 fusion gene, preferably a fusion gene having a sequence at least 70% identical to the sequence of SOS1-ADCY3 (SEQ ID NO. 9).

The phosphorylation status PKA and/or CREB may be determined by methods known by the skilled artisan. Such methods include antibody based assays using phopho-specific antibodies. In a preferred embodiment the phosphorylation status is determined by immunohistochemistry (IHC).

It will be apparent that the methods and components of the present invention, the proteins obtained thereby, as well as the uses as substantially described herein or illustrated in the description and the examples, are also subject of the present invention and claimed herewith. In this respect, it is also understood that the embodiments as described in the description and/or any one of the examples, can be independently used and combined with any one of the embodiments described hereinbefore and claimed in the appended claims set. Thus, these and other embodiments are disclosed and encompassed by the description and examples of the present invention.

In particular the invention relates to the following aspects:

In a first aspect the invention relates to a method for diagnosing lung cancer in a patient comprising the steps of: - determining the presence and/or expression of at least one fusion gene in a sample derived from said patient; and - attributing the presence or expression of said

fusion gene to the presence of lung cancer in said patient; wherein the fusion gene comprises one or more nucleotide sequences encoding a protein or a fragment of the protein, said protein being selected from the group consisting of a guanine nucleotide exchange factor for RAS proteins, an adenylyl cyclase, a Zink finger protein, a protein tyrosine kinase, a member of
5 ADP-ribosylation factor GTPase activating (ARF-GAP) protein family, a member of the IGF-1R pathway, a polypeptide hormone precursor, an adducin, a fibronectin type III domain containing protein, a germinal center kinase (GCKs), an chromosome open reading frame, a member of the Ras-association domain family (RASSF), a member of the E2F family of transcription factors, a ribosomal protein of the L14E family, a member of CDS1 subfamily of
10 serine/threonine protein kinases, a THO complex, a member of P-type lectin family, a membrane glycoprotein of T-Lymphocytes, a Z-DNA binding protein, a breast carcinoma amplified sequence, an early B-cell factor, a class V ATPase, a protein involved in glycosylphosphatidylinositol (GPI)-anchor biosynthesis, a member of the ESCRT-III protein complex, a repressor of glucocorticoid receptor, a discslarge (Drosophila) homolog-associated
15 protein, a G-protein coupled receptor, and a cholesterol ester hydrolase; and wherein the fusion gene may comprise one or more nucleotide sequences from a non-protein coding RNA or a fragment thereof.

In a second aspect the invention relates to the method according to the first aspect, wherein said
20 guanine nucleotide exchange factor for RAS proteins is SOS1 (SEQ ID NO. 2) or a protein having at least 70% identity thereto;

wherein said adenylyl cyclase is selected from the group of adenylyl cyclases consisting of ADCY1 (SEQ ID NO. 141), ADCY2 (SEQ ID NO. 142), ADCY3 (SEQ ID NO. 6), ADCY4 (SEQ ID NO. 143), ADCY5 (SEQ ID NO. 144), ADCY6 (SEQ ID NO. 145), ADCY7 (SEQ
25 ID NO. 146), ADCY8 (SEQ ID NO. 147), ADCY9 (SEQ ID NO. 148), and ADCY10 (SEQ ID NO. 149); or a protein having at least 70% identity thereto;

wherein said Zink finger protein is ZNF142 (SEQ ID NO. 12) or a protein having at least 70% identity thereto;

wherein said protein tyrosine kinase is PTK6 (SEQ ID NO. 16) or a protein having at least 70%
30 identity thereto;

wherein said member of ADP-ribosylation factor GTPase activating (ARF-GAP) protein family is AGAP1 (SEQ ID NO. 22) or a protein having at least 70% identity thereto;

wherein said member of the IGF-1R pathway is IGFBP2 (SEQ ID NO. 26) or a protein having at least 70% identity thereto;

wherein said polypeptide hormone precursor is POMC (SEQ ID NO. 32) or a protein having at least 70% identity thereto;

5 wherein said adducin is ADD1 (SEQ ID NO. 36) or a protein having at least 70% identity thereto;

wherein said fibronectin type III domain containing protein is FNDC3B (SEQ ID NO. 41) or a protein having at least 70% identity thereto;

10 wherein said germinal center kinase (GCK) is TNIK (SEQ ID NO. 45) or a protein having at least 70% identity thereto;

wherein said chromosome open reading frame is C12orf11 (SEQ ID NO. 51) or a protein having at least 70% identity thereto;

wherein said member of the Ras-association domain family (RASSF) is RASSF8 (SEQ ID NO. 55) or a protein having at least 70% identity thereto;

15 wherein said member of the E2F family of transcription factors is selected from the group consisting of E2F4 (SEQ ID NO. 61) and E2F3 (SEQ ID NO. 125), or a protein having at least 70% identity thereto;

wherein said ribosomal protein of the L14E family is RPL14 (SEQ ID NO. 65) or a protein having at least 70% identity thereto;

20 wherein said member of CDS1 subfamily of serine/threonine protein kinases is CHEK2 (SEQ ID NO. 71) or a protein having at least 70% identity thereto;

wherein said THO complex is THOC5 (SEQ ID NO. 75) or a protein having at least 70% identity thereto;

25 wherein said member of P-type lectin family is M6PR (SEQ ID NO. 80) or a protein having at least 70% identity thereto;

wherein said membrane glycoprotein of T-Lymphocytes is CD4 (SEQ ID NO. 84) or a protein having at least 70% identity thereto; and

wherein said a Z-DNA binding protein is ZBP1 (SEQ ID NO. 89) or a protein having at least 70% identity thereto;

30 wherein said breast carcinoma amplified sequence is BCAS1 (SEQ ID NO. 93) or a protein having at least 70% identity thereto;

wherein said early B-cell factor is EBF1 (SEQ ID NO. 98) or a protein having at least 70%

identity thereto;

wherein said class V ATPase is ATP10B (SEQ ID NO. 102) or a protein having at least 70% identity thereto;

wherein said protein involved in glycosylphosphatidylinositol (GPI)-anchor biosynthesis is PIGF (SEQ ID NO. 107) or a protein having at least 70% identity thereto;

wherein said member of the ESCRT-III protein complex is CHMP3 (SEQ ID NO. 111) or a protein having at least 70% identity thereto;

wherein said repressor of glucocorticoid receptor is GRLF1 (SEQ ID NO. 116) or a protein having at least 70% identity thereto;

wherein said discs, large (*Drosophila*) homolog-associated protein is DLGAP1 (SEQ ID NO. 120) or a protein having at least 70% identity thereto

wherein said G-protein coupled receptor is GPR160 (SEQ ID NO. 133) or a protein having at least 70% identity thereto;

wherein said cholesterol ester hydrolase is NCEH1 (SEQ ID NO. 137) or a protein having at least 70% identity thereto; and

wherein said non-protein coding RNA is FLJ22536 (SEQ ID. 128).

In a third aspect the invention relates to the method according to the first or the second aspect, wherein said fusion gene comprises one or more nucleotide sequence of a gene or non-protein coding RNA; or a fragment of the gene or non-protein coding RNA, said gene or non-coding RNA having a sequence with at least 70% identity to a sequence selected from the group consisting of SOS1 (SEQ ID NO. 1), ADCY3 (SEQ ID NO. 5), ZNF142 (SEQ ID NO. 11), PTK6 (SEQ ID NO. 15), AGAP1 (SEQ ID NO. 21), IGFBP2 (SEQ ID NO. 25), POMC (SEQ ID NO. 31), ADD1 (SEQ ID NO. 35), FNDC3B (SEQ ID NO. 40), TNIK (SEQ ID NO. 44), C12orf11 (SEQ ID NO. 50), RASSF8 (SEQ ID NO. 54), E2F4 (SEQ ID NO. 60), RPL14 (SEQ ID NO. 64), CHEK2 (SEQ ID NO. 70), THOC5 (SEQ ID NO. 74), M6PR (SEQ ID NO. 79), CD4 (SEQ ID NO. 83), ZBP1 (SEQ ID NO. 88), BCAS1 (SEQ ID NO. 92), EBF1 (SEQ ID NO. 97), ATP10B (SEQ ID NO. 101), PIGF (SEQ ID NO. 106), CHMP3 (SEQ ID NO. 110), GRLF1 (SEQ ID NO. 115), DLGAP1 (SEQ ID NO. 119), E2F3 (SEQ ID NO. 124), GPR160 (SEQ ID NO. 132), NCEH1 (SEQ ID NO. 136); and FLJ22536 (SEQ ID NO. 128).

In a fourth aspect the invention relates to the method according to aspect 3, wherein said fusion

gene comprises one or more nucleotide sequences of a fragment of the gene or the non-protein coding RNA, said fragment having a sequence with at least 70% identity to a sequence selected from the group consisting of SOS1 (SEQ ID NO. 3), ADCY3 (SEQ ID NO. 7), ZNF142 (SEQ ID NO. 13), PTK6 (SEQ ID NO. 17), AGAP1 (SEQ ID NO. 23), IGFBP2 (SEQ ID NO. 27),
5 POMC (SEQ ID NO. 33), ADD1 (SEQ ID NO. 37), FNDC3B (SEQ ID NO. 42), TNIK (SEQ ID NO. 46), C12orf11 (SEQ ID NO. 52), RASSF8 (SEQ ID NO. 56), E2F4 (SEQ ID NO. 62), RPL14 (SEQ ID NO. 66), CHEK2 (SEQ ID NO. 72), THOC5 (SEQ ID NO. 76), M6PR (SEQ ID NO. 81), CD4 (SEQ ID NO. 85), ZBP1 (SEQ ID NO. 90), BCAS1 (SEQ ID NO. 94), EBF1 (SEQ ID NO. 99), ATP10B (SEQ ID NO. 103), PIGF (SEQ ID NO. 108), CHMP3 (SEQ ID NO. 112), GRLF1 (SEQ ID NO. 117), DLGAP1 (SEQ ID NO. 121), E2F3 (SEQ ID NO. 126),
10 FLJ22536 (SEQ ID NO. 129), GPR160 (SEQ ID NO. 134), and NCEH1 (SEQ ID NO. 138).

In a fifth aspect the invention relates to the method according to any one of the first to the fourth aspect, wherein the fusion gene is selected from the group of fusion genes consisting of
15 SOS1-ADCY3, ZNF142-PTK6, AGAP1-IGFBP2, POMC-ADD1, FNDC3B-TNIK, C12orf11-RASSF8, E2F4-RPL14, CHEK2-THOC5, M6PR-CD4, ZBP1-BCAS1, EBF1-ATP10B, PIGF-CHMP3, GRLF1-DLGAP1, E2F3-FLJ22536, and GPR160-NCEH1.

In a sixth aspect the invention relates to the method according to the fifth aspect, wherein the
20 fusion gene has a sequence with at least 70% identity to a sequence selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 9), ZNF142-PTK6 (SEQ ID NO. 19), AGAP1-IGFBP2 (SEQ ID NO. 29), POMC-ADD1 (SEQ ID NO. 39), FNDC3B-TNIK (SEQ ID NO. 48), C12orf11-RASSF8 (SEQ ID NO. 58), E2F4-RPL14 (SEQ ID NO. 68), CHEK2-THOC5 (SEQ ID NO. 77), M6PR-CD4 (SEQ ID NO. 86), ZBP1-BCAS1 (SEQ ID NO. 95), EBF1-
25 ATP10B (SEQ ID NO. 104), PIGF-CHMP3 (SEQ ID NO. 113), GRLF1-DLGAP1 (SEQ ID NO. 122), E2F3-FLJ22536 (SEQ ID NO. 130), and GPR160-NCEH1 (SEQ ID NO. 139).

In a seventh aspect the invention relates to the method according to any one of the first to the sixth aspect, wherein the lung cancer is selected from a group consisting of any of the lung
30 cancer tumor types as classified in the WHO classification of lung tumors (2004), preferably selected from the group consisting of squamous, adenocarcinoma, small cell carcinoma, large cell carcinoma (including large cell neuroendocrine carcinoma), and carcinoid tumor.

In an eighth aspect the invention relates to the method according to any one of the first to the seventh aspect, wherein the presence and/or expression of said fusion gene is detected by a hybridisation and/or an amplification assay.

5

In a ninth aspect the invention relates to the method according to the eighth aspect, wherein the hybridisation and/or amplification assay is selected from the group consisting of polymerase chain reaction (PCR), real-time PCR, reverse transcriptase PCR (RT-PCR), fluorescent *in situ* hybridisation (FISH), chromogenic *in situ* hybridisation (CISH), break-apart *in situ* hybridisation (ba-FISH), and silver *in situ* hybridisation (SISH).

10

In a tenth aspect the invention relates to a method of selecting a cell, tissue or cell culture with susceptibility to a MEK inhibitor, comprising the steps: - determining the presence and/or expression of a fusion gene in said cell, tissue or cell culture, wherein the fusion gene comprises a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase in said cell, tissue or cell culture; and - attributing the presence and/or expression of said fusion gene to the susceptibility of the cell, tissue or cell culture to a MEK inhibitor.

15

in an eleventh aspect the invention relates to a method for identifying a responder or non-responder to a MEK inhibitor, the method comprising the steps: - providing a sample from a patient suspected to suffer or being prone to suffer from lung cancer; and - determining the presence and/or expression of a fusion gene comprising a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase, or a fragment thereof; wherein the presence and/or expression of said fusion gene is indicative for a responder to a MEK inhibitor.

20

25

In a twelfth aspect the invention relates to a method of determining resistance of a cell, tissue or cell culture to an ALK inhibitor, comprising the steps: - determining the presence and/or expression level of a fusion gene in said cell, tissue or cell culture, wherein the fusion gene comprises a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase in said cell, tissue or cell culture; - comparing the determined

30

expression level to the expression level in a control cell, tissue or cell culture expressing said fusion gene and not being resistant to an ALK inhibitor; and - attributing a determined expression level higher than in the control cell, tissue or cell culture to the resistance of the cell, tissue or cell culture to a ALK inhibitor; and/or - attributing a determined expression equal or lower than in the control cell, tissue or cell culture to the susceptibility of the cell, tissue or cell culture to a ALK inhibitor.

In a thirteenth aspect the invention relates to a method for identifying a responder or non-responder to an ALK inhibitor, the method comprising the steps: - determining the presence/expression level of a fusion gene in a sample from a patient suspected to suffer or being prone to suffer from lung cancer, said fusion gene comprising a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase, or a fragment thereof; and - comparing the determined expression level to the expression level in a sample from a control patient, wherein the control patient is a responder to ALK inhibitors and expresses said fusion gene; wherein a determined expression higher than in said sample of a control patient is indicative for a non-responder to a ALK inhibitor; and wherein a determined expression equal or lower than in said sample of a control patient is indicative for a responder to a ALK inhibitor.

In a fourteenth aspect the invention relates to a method of selecting a cell, tissue or cell culture with susceptibility to a HAT inhibitor, comprising the steps: - determining the presence and/or expression of a fusion gene in said cell, tissue or cell culture, wherein the fusion gene comprises a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase in said cell, tissue or cell culture; and - attributing the presence and/or expression of said fusion gene to the susceptibility of the cell, tissue or cell culture to a HAT inhibitor.

In a fifteenth aspect the invention relates to a method for identifying a responder or non-responder to a HAT inhibitor, the method comprising the steps: - determining the presence and/or expression of a fusion gene in a sample from a patient suspected to suffer or being prone to suffer from lung cancer, said fusion gene comprising a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase, or a fragment

thereof; wherein the presence and/or expression of said fusion gene is indicative for a responder to a HAT inhibitor.

5 In a sixteenth aspect the invention relates to a method of selecting a cell, tissue or cell culture with susceptibility to a PKA inhibitor, comprising the steps: - determining the presence and/or expression of a fusion gene in said cell, tissue or cell culture, wherein the fusion gene comprises a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase in said cell, tissue or cell culture; and -attributing the presence and/or expression of said fusion gene to the susceptibility of the cell, tissue or cell culture to a
10 PKA inhibitor.

In a seventeenth aspect the invention relates to a method for identifying a responder or non-responder to a PKA inhibitor, the method comprising the steps: - determining the presence and/or expression of a fusion gene in a sample from a patient suspected to suffer or being prone
15 to suffer from lung cancer, said fusion gene comprising a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase, or a fragment thereof; wherein the presence and/or expression of said fusion gene is indicative for a responder to a PKA inhibitor.

20 In an eighteenth aspect the invention relates to a method according to any one of tenth to the seventeenth aspect, wherein said adenylyl cyclase is selected from the group consisting of ADCY1 (SEQ ID NO. 141), ADCY2 (SEQ ID NO. 142), ADCY3 (SEQ ID NO. 6), ADCY4 (SEQ ID NO. 143), ADCY5 (SEQ ID NO. 144), ADCY6 (SEQ ID NO. 145), ADCY7 (SEQ ID NO. 146), ADCY8 (SEQ ID NO. 147), ADCY9 (SEQ ID NO. 148), and ADCY10 (SEQ ID
25 NO. 149); or a protein having at least 70% identity thereto.

In a nineteenth aspect the invention relates to a method according to any one of the tenth to the eighteenth aspect, wherein said fusion gene further comprises a nucleotide sequence encoding a second protein or a fragment of the second protein, said second protein being a guanine
30 nucleotide exchange factor for RAS proteins.

In a twentieth aspect the invention relates to a method according to the nineteenth aspect,

wherein said guanine nucleotide exchange factor for RAS proteins is SOS1 (SEQ ID NO. 2) or a protein having at least 70% identity thereto.

In a twenty-first aspect the invention relates to method according to any one of the tenth to the twentieth aspect, wherein the fusion gene comprises one or more nucleotide sequence of a gene or a fragment of the gene, said gene having at least 70% identity to a sequence selected from the group consisting of SOS1 (SEQ ID NO. 1), and ADCY3 (SEQ ID NO. 5), or a fragment thereof.

10 In a twenty-second aspect the invention relates to a method according to twenty-first aspect, wherein the fusion gene comprises one or more nucleotide sequence of a fragment of the gene, said fragment of the gene being selected from the group consisting of SOS1 (SEQ ID NO. 3), and ADCY3 (SEQ ID NO. 7).

15 In a twenty-third aspect the invention relates to the method according to any one of the tenth to the twenty-second aspect, wherein the fusion gene is a SOS1-ADCY3 fusion gene.

In a twenty-fourth aspect the invention relates to the method according to any one of the tenth to the twenty-third aspect, wherein the fusion gene has a sequence at least 70% identical to SOS1-ADCY3 (SEQ ID NO. 9).

20 In a twenty-fifth aspect the invention relates to a method of selecting a cell, tissue or cell culture with susceptibility to an FGFR1 inhibitor, comprising the steps: - determining the presence and/or expression of a fusion gene in said cell, tissue or cell culture; wherein said fusion gene comprises one or more nucleotide sequence encoding a protein or a fragment of the protein, said protein being selected from the group consisting of a member of the E2F family of transcription factors, a ribosomal protein of the L14E family; and – attributing the presence and/or expression of said fusion gene to susceptibility of the cell, tissue or cell culture to a FGFR1 inhibitor.

30 In a twenty-sixth aspect the invention relates to a method for identifying a responder or non-responder to an FGFR1 inhibitor, the method comprising the steps: - determining the presence

and/or expression of a fusion gene in a sample from a patient suspected to suffer or being prone to suffer from cancer, said fusion gene comprising a nucleotide sequence encoding a protein or a fragment of the protein, said protein selected from the group consisting of a member of the E2F family of transcription factors, and a ribosomal protein of the L14E family; wherein the presence and/or expression of said fusion gene is indicative for a responder to a FGFR1 inhibitor.

In a twenty-seventh aspect the invention relates to the method according to the twenty-fifth or twenty-sixth aspect, wherein said member of the E2F family of transcription factors is E2F4 (SEQ ID NO. 61); or a protein having at least 70% identity thereto; wherein said ribosomal protein of the L14E family is RPL14 (SEQ ID NO. 65); or a protein having at least 70% identity thereto.

In a twenty-eighth aspect the invention relates to the method according to any one of the twenty-fifth to the twenty-seventh aspect, wherein said fusion gene comprises one or more nucleotide sequence of a gene or a fragment of the gene, said gene having a sequence with at least 70% identity to a sequence selected from the group consisting of E2F4 (SEQ ID NO. 60), and RPL14 (SEQ ID NO. 64).

In a twenty-ninth aspect the invention relates to the method according to the twenty-eighth aspect, wherein said fusion gene comprises one or more nucleotide sequence of a fragment of the gene, said fragment having a sequence with at least 70% identity to a sequence selected from the group consisting of E2F4 (SEQ ID NO. 62), and RPL14 (SEQ ID NO. 66).

In a thirtieth aspect the invention relates to the method according to any one of the twenty-fifth to the twenty-ninth aspect, wherein the fusion gene is a E2F4-RPL14 fusion gene.

In a thirty-first aspect the invention relates to the method according to the thirtieth aspect, wherein the fusion gene has a nucleotide sequence having at least 70% identity to the sequence of E2F4-RPL14 (SEQ ID NO. 68).

In a thirty-second aspect the invention relates to a method of selecting a cell, tissue or cell

culture with susceptibility to a PTK inhibitor, comprising the steps: - determining the presence and/or expression of a fusion gene in said cell, tissue or cell culture, wherein the fusion gene comprises a nucleotide sequence encoding a protein or a fragment of the protein, said protein being a protein tyrosine kinase, in said cell, tissue or cell culture; and - attributing the presence and/or expression of said fusion gene to the susceptibility of the cell, tissue or cell culture to a PTK inhibitor.

In a thirty-third aspect the invention relates to a method for identifying a responder or non-responder to a PTK inhibitor, the method comprising the steps: - determining the presence and/or expression of a fusion gene in a sample from a patient suspected to suffer or being prone to suffer from lung cancer, said fusion gene comprising a nucleotide sequence encoding a protein or a fragment of the protein, said protein being a protein tyrosine kinase, or a fragment thereof; wherein the presence and/or expression of said fusion gene is indicative for a responder to a PTK inhibitor.

In a thirty-fourth aspect the invention relates to the method according to the thirty-second or thirty-third aspect, wherein said protein tyrosine kinase is PTK6 (SEQ ID NO. 16) or a protein having at least 70% identity thereto; and the PTK inhibitor is a PTK6 inhibitor.

In a thirty-fifth aspect the invention relates to the method according to any one of the thirty-second to the thirty-fourth aspect, wherein the fusion gene further comprises nucleotide sequence encoding a protein or a fragment of the protein, said protein being a Zinc finger protein.

In a thirty-sixth aspect the invention relates to the method according to the thirty-fifth aspect, wherein said Zinc finger protein is ZNF142 (SEQ ID NO. 12) or a protein having at least 70% identity thereto.

In a thirty-seventh aspect the present invention relates to the method according to any one of the thirty-second to the thirty-sixth aspect, wherein the fusion gene encodes a protein sequence of ZNF142-PTK6 (SEQ ID NO. 20) or a protein having at least 70% identity thereto.

In a thirty-eighth aspect the invention relates to a method of selecting a cell, tissue or cell culture with susceptibility to a c-Jun inhibitor, comprising the steps: - determining the presence and/or expression of a fusion gene in said cell, tissue or cell culture, wherein the fusion gene comprises a nucleotide sequence encoding a protein or a fragment of the protein, said protein being a germinal center kinase, in said cell, tissue or cell culture; and - attributing the presence and/or expression of said fusion gene to the susceptibility of the cell, tissue or cell culture to a c-Jun inhibitor.

In a thirty-ninth aspect the invention relates to a method for identifying a responder or non-responder to a c-Jun inhibitor, the method comprising the steps: - determining the presence and/or expression of a fusion gene in a sample from a patient suspected to suffer or being prone to suffer from lung cancer, said fusion gene comprising a nucleotide sequence encoding a protein or a fragment of the protein, said protein being germinal center kinase, or a fragment thereof; wherein the presence and/or expression of said fusion gene is indicative for a responder to a c-Jun inhibitor.

In a fortieth aspect the invention relates to the method according to the thirty-eighth or thirty-ninth aspect, wherein said germinal center kinase is TNIK (SEQ ID NO. 45) or a protein having at least 70% identity thereto.

In a forty-first aspect the invention relates to the method according to any one of the thirty-eighth to the fortieth aspect, wherein the fusion gene further comprises nucleotide sequence encoding a protein or a fragment of the protein, said protein being fibronectin type III domain containing protein.

In a forty-second aspect the invention relates to the method according to the forty first aspect, wherein said fibronectin type III domain containing protein is FNDC3B (SEQ ID NO. 41) or a protein having at least 70% identity thereto.

In a forty-third aspect the invention relates to the method according to any one of the thirty-eighth to the forty-second aspect, wherein said fusion gene encodes a protein sequence of FNDC3B-TNIK (SEQ ID NO 49) or a protein having at least 70% identity thereto.

In a forty-fourth aspect the invention relates to a method for identifying compounds for the treatment of lung cancer comprising the steps: - providing a cell expressing a fusion gene, wherein the fusion gene comprises one or more nucleotide sequence encoding a protein or a fragment of the protein, said protein being selected from the group consisting of a guanine nucleotide exchange factor for RAS proteins, an adenylyl cyclase, a Zink finger protein, a protein tyrosine kinase, a member of ADP-ribosylation factor GTPase activating (ARF-GAP) protein family, a member of the IGF-1R pathway, a polypeptide hormone precursors, an adducin, a fibronectin type III domain containing protein, a germinal center kinase (GCKs), a chromosome open reading frame, a member of the Ras-association domain family (RASSF), a member of the E2F family of transcription factors, a ribosomal protein of the L14E family; - contacting said cell to a candidate compound; and - evaluating the viability of said cell contacted with said candidate compound.

In a forty-fifth aspect the invention relates to the method according to the forty-fourth aspect, wherein the cell is a cell of a cell line selected from the group consisting of lung cancer cell-lines.

In a forty-sixth aspect the invention relates to the method according to the forty-fourth or forty-fifth aspect, wherein said guanine nucleotide exchange factor for RAS proteins is SOS1 (SEQ ID NO. 2) or a protein having at least 70% identity thereto; wherein said adenylyl cyclase is selected from the group of adenylyl cyclises consisting of ADCY1 (SEQ ID NO. 141), ADCY2 (SEQ ID NO. 142), ADCY3 (SEQ ID NO. 6) ADCY4 (SEQ ID NO. 143), ADCY5 (SEQ ID NO. 144), ADCY6 (SEQ ID NO. 145), ADCY7 (SEQ ID NO. 146), ADCY8 (SEQ ID NO. 147), ADCY9 (SEQ ID NO. 148), and ADCY10 (SEQ ID NO. 149); or a protein having at least 70% identity thereto; wherein said Zink finger protein is ZNF142 (SEQ ID NO. 12) or a protein having at least 70% identity thereto; wherein said protein tyrosine kinase is PTK6 (SEQ ID NO. 16) or a protein having at least 70% identity thereto; wherein said member of ADP-ribosylation factor GTPase activating (ARF-GAP) protein family is AGAP1 (SEQ ID NO. 22) or a protein having at least 70% identity thereto; wherein said member of the IGF-1R pathway is IGFBP2 (SEQ ID NO. 26) or a protein having

at least 70% identity thereto;

wherein said polypeptide hormone precursor is POMC (SEQ ID NO. 32) or a protein having at least 70% identity thereto;

wherein said adducin is ADD1 (SEQ ID NO. 36) or a protein having at least 70% identity thereto;

wherein said fibronectin type III domain containing protein is FNDC3B (SEQ ID NO. 41) or a protein having at least 70% identity thereto;

wherein said germinal center kinase (GCK) is TNIK (SEQ ID NO. 45) or a protein having at least 70% identity thereto;

wherein chromosome open reading frame is C12orf11 (SEQ ID NO. 51) or a protein having at least 70% identity thereto;

wherein said member of the Ras-association domain family (RASSF) is RASSF8 (SEQ ID NO. 55) or a protein having at least 70% identity thereto;

wherein said member of the E2F family of transcription factors is E2F4 (SEQ ID NO. 61), or a protein having at least 70% identity thereto; and

wherein said ribosomal protein of the L14E family is RPL14 (SEQ ID NO. 65) or a protein having at least 70% identity thereto.

In a forty-seventh aspect the invention relates to the method according to any one of the forty-fourth to the forty-sixth aspect, wherein said fusion gene comprises one or more nucleotide sequence of a gene or a fragment of the gene, said gene having a sequence at least 70% identity to a sequence selected from the group consisting of SOS1 (SEQ ID NO. 1), ADCY3 (SEQ ID NO. 5), ZNF142 (SEQ ID NO. 11), PTK6 (SEQ ID NO. 15), AGAP1 (SEQ ID NO. 21), IGFBP2 (SEQ ID NO. 25), POMC (SEQ ID NO. 31), ADD1 (SEQ ID NO. 35), FNDC3B (SEQ ID NO. 40), TNIK (SEQ ID NO. 44), C12orf11 (SEQ ID NO. 50), RASSF8 (SEQ ID NO. 54), E2F4 (SEQ ID NO. 60), and RPL14 (SEQ ID NO. 64).

In a forty-eighth aspect the invention relates to the method according to the forty-seventh aspect, wherein said fusion gene comprises one or more nucleotide sequence of a fragment of the gene, said fragment having a sequence at least 70% identity to a sequence selected from the group consisting of SOS1 (SEQ ID NO. 3), ADCY3 (SEQ ID NO. 7), ZNF142 (SEQ ID NO. 13), PTK6 (SEQ ID NO. 17), AGAP1 (SEQ ID NO. 23), IGFBP2 (SEQ ID NO. 27), POMC (SEQ

ID NO. 33), ADD1 (SEQ ID NO. 37), FNDC3B (SEQ ID NO. 42), TNIK (SEQ ID NO. 46), C12orf11 (SEQ ID NO. 52), RASSF8 (SEQ ID NO. 56), E2F4 (SEQ ID NO. 62), and RPL14 (SEQ ID NO. 66).

- 5 In a forty-ninth aspect the invention relates to the method according to any one of the forty-fourth to the forty-eighth aspect, wherein said fusion gene is a fusion gene selected from the group of fusion genes consisting of SOS1-ADCY3, ZNF142-PTK6, AGAP1-IGFBP2, POMC-ADD1, FNDC3B-TNIK, C12orf11-RASSF8, and E2F4-RPL14.
- 10 In a fiftieth aspect the invention relates to the method according to any one of claims the forty-fourth to the forty-ninth aspect, wherein the fusion gene has a sequence with at least 70% identity to a sequence selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 9), ZNF142-PTK6 (SEQ ID NO. 19), AGAP1-IGFBP2 (SEQ ID NO. 29), POMC-ADD1 (SEQ ID NO. 39), FNDC3B-TNIK (SEQ ID NO. 48), C12orf11-RASSF8 (SEQ ID NO. 58), and E2F4-
- 15 RPL14 (SEQ ID NO. 68).

In a fifty-first aspect the invention relates to the use of a cell line expressing at least one of the fusion genes selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 9), ZNF142-PTK6 (SEQ ID NO. 19), AGAP1-IGFBP2 (SEQ ID NO. 29), POMC-ADD1 (SEQ ID NO. 39),

20 FNDC3B-TNIK (SEQ ID NO. 48), C12orf11-RASSF8 (SEQ ID NO. 58), and E2F4-RPL14 (SEQ ID NO. 68) for screening and/or validation of a medicament for the treatment of lung cancer.

In a fifty-second aspect the invention relates to the use according to the fifty-first aspect,

25 wherein the cell is a transgenic cell.

In a fifty-third aspect the invention relates to the use according to the fifty-first or fifty-second aspect, wherein the lung cancer is selected from the group of lung cancer types consisting of squamous, adenocarcinoma, small cell carcinoma, large cell carcinoma (including large cell

30 neuroendocrine carcinoma), and carcinoid tumor.

In a fifty-fourth aspect the invention relates to a nucleic acid comprising a sequence at least

70% identical to a sequence selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 9), ZNF142-PTK6 (SEQ ID NO. 19), AGAP1-IGFBP2 (SEQ ID NO. 29), POMC-ADD1 (SEQ ID NO. 39), FNDC3B-TNIK (SEQ ID NO. 48), C12orf11-RASSF8 (SEQ ID NO. 58), E2F4-RPL14 (SEQ ID NO. 68), CHEK2-THOC5 (SEQ ID NO. 77), M6PR-CD4 (SEQ ID NO. 86),
5 ZBP1-BCAS1 (SEQ ID NO. 95), EBF1-ATP10B (SEQ ID NO. 104), PIGF-CHMP3 (SEQ ID NO. 113), GRLF1-DLGAP1 (SEQ ID NO. 122), E2F3-FLJ22536 (SEQ ID NO. 130), and GPR160-NCEH1 (SEQ ID NO. 139).

In a fifty-fifth aspect the invention relates to a nucleic acid encoding a sequence at least 70% identical to a protein sequence selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 10), ZNF142-PTK6 (SEQ ID NO. 20), AGAP1-IGFBP2 (SEQ ID NO. 30), FNDC3B-TNIK (SEQ ID NO. 49), C12orf11-RASSF8 (SEQ ID NO. 59), E2F4-RPL14 (SEQ ID NO. 69), CHEK2-THOC5 (SEQ ID NO. 78), M6PR-CD4 (SEQ ID NO. 87), ZBP1-BCAS1 (SEQ ID NO. 96), EBF1-ATP10B (SEQ ID NO. 105), GRLF1-DLGAP1 (SEQ ID NO. 123), and
15 E2F3-FLJ22536 (SEQ ID NO. 131).

In a fifty-sixth aspect the invention relates to a protein having a sequence with at least 70 % identity to a sequence selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 10), ZNF142-PTK6 (SEQ ID NO. 20), AGAP1-IGFBP2 (SEQ ID NO. 30), FNDC3B-TNIK (SEQ ID NO. 49), C12orf11-RASSF8 (SEQ ID NO. 59), E2F4-RPL14 (SEQ ID NO. 69), CHEK2-THOC5 (SEQ ID NO. 78), M6PR-CD4 (SEQ ID NO. 87), ZBP1-BCAS1 (SEQ ID NO. 96), EBF1-ATP10B (SEQ ID NO. 105), GRLF1-DLGAP1 (SEQ ID NO. 123), and E2F3-FLJ22536 (SEQ ID NO. 131).

25 In a fifty-seventh aspect the invention relates to a kit useful for carrying out the method of any one of the first to the thirty-first aspect, said kit comprising an oligonucleotide or polynucleotide capable of amplifying and/or specifically detecting a fusion gene selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 9), ZNF142-PTK6 (SEQ ID NO. 19), AGAP1-IGFBP2 (SEQ ID NO. 29), POMC-ADD1 (SEQ ID NO. 39), FNDC3B-TNIK (SEQ ID NO. 48), C12orf11-RASSF8 (SEQ ID NO. 58), E2F4-RPL14 (SEQ ID NO. 68), CHEK2-THOC5 (SEQ ID NO. 77), M6PR-CD4 (SEQ ID NO. 86), ZBP1-BCAS1 (SEQ ID NO. 95), EBF1-ATP10B (SEQ ID NO. 104), PIGF-CHMP3 (SEQ ID NO. 113), GRLF1-DLGAP1 (SEQ
30

ID NO. 122), E2F3-FLJ22536 (SEQ ID NO. 130), and GPR160-NCEH1 (SEQ ID NO. 139); and/or said kit comprising an antibody specifically binding to a fusion protein selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 10), ZNF142-PTK6 (SEQ ID NO. 20), AGAP1-IGFBP2 (SEQ ID NO. 30), FNDC3B-TNIK (SEQ ID NO. 49), C12orf11-RASSF8 (SEQ ID NO. 59), E2F4-RPL14 (SEQ ID NO. 69), CHEK2-THOC5 (SEQ ID NO. 78), M6PR-CD4 (SEQ ID NO. 87), ZBP1-BCAS1 (SEQ ID NO. 96), EBF1-ATP10B (SEQ ID NO. 105), GRLF1-DLGAP1 (SEQ ID NO. 123), and E2F3-FLJ22536 (SEQ ID NO. 131).

In a fifty-eighth aspect the invention relates to a method for identifying the presence and expression of a fusion gene comprising a nucleic acid encoding an adenylyl cyclase in a sample, the method comprising the step of: (i) determining the phosphorylation status of PKA and/or CREB; (ii) comparing the determined phosphorylation status to the phosphorylation status of PKA and/or CREB in a control; and (iii) attributing a higher phosphorylation status of PKA and/or CREB in determined step (i) than in the control to the presence and expression of a fusion gene comprising a nucleotide sequence encoding a protein, said protein being an adenylyl cyclase.

In a fifty-ninth aspect the invention relates to the method according to the fifty-eighth aspect, wherein said adenylyl cyclase is selected from the group consisting of ADCY1 (SEQ ID NO. 141), ADCY2 (SEQ ID NO. 142), ADCY3 (SEQ ID NO. 6) ADCY4 (SEQ ID NO. 143), ADCY5 (SEQ ID NO. 144), ADCY6 (SEQ ID NO. 145), ADCY7 (SEQ ID NO. 146), ADCY8 (SEQ ID NO. 147), ADCY9 (SEQ ID NO. 148), and ADCY10 (SEQ ID NO. 149); or a protein having at least 70% identity thereto.

In a sixtieth aspect the invention relates to the method according to the fifty-eighth or fifty-ninth aspect, wherein a higher phosphorylation status of PKA and/or CREB determined in step (i) than in the control is attributed to the presence and expression of a fusion gene further comprising a nucleotide sequence encoding a second protein or a fragment of the second protein, said second protein being a guanine nucleotide exchange factor for RAS proteins.

In a sixty-first aspect the invention relates to the method according to the sixtieth aspect, wherein said second protein being a guanine nucleotide exchange factor for RAS proteins is

proteins is SOS1 (SEQ ID NO. 2) or a protein having at least 70% identity thereto.

In a sixty-second aspect the invention relates to the method according to the sixtieth or sixty-first aspect, wherein the fusion gene is a SOS1-ADCY3 fusion gene, preferably a fusion gene
5 having a sequence at least 70% identical to the sequence of SOS1-ADCY3 (SEQ ID NO. 9).

In a sixty-third aspect the invention relates to the method according to any one of fifty-eighth to the sixty-second aspect, wherein the phosphorylation status is determined by fluorescent *in situ* hybridisation (FISH).

10 The present invention is further illustrated by the following examples and figures.

EXAMPLES

15 1. Identification of fusion genes and RNA-seq

RNAseq was performed on cDNA libraries prepared from PolyA+ RNA extracted from tumor cells. The average insert size was 250 bp to allow for paired-end reads of 2x95bp without any overlap. RNAseq affords robust detection of expressed fusion transcripts in addition to
20 enabling integrated analyses of orthogonal genomics datasets by providing gene expression data.

For the analysis of RNAseq data, software was used that afford accurate and efficient mapping and downstream analysis of transcribed genes. Candidates for fusion genes were validated
25 when at least 1 read-pair mapped uniquely within the human genome to two different genes (hereinafter referred to as encompassing fragments), and at least one 95-bp read unambiguously spanned a junction between two exons of different genes (hereinafter referred to as fusion-spanning reads), and when the candidate was found with at least 5x coverage.

30 Identified fusion genes are SOS1-ADCY3 (SEQ ID NO. 9), ZNF142-PTK6 (SEQ ID NO. 19), AGAP1-IGFBP2 (SEQ ID NO. 29), POMC-ADD1 (SEQ ID NO. 39), FNDC3B-TNIIK (SEQ ID NO. 48), C12orf11-RASSF8 (SEQ ID NO. 58), E2F4-RPL14 (SEQ ID NO. 68), CHEK2-

THOC5 (SEQ ID NO. 77), M6PR-CD4 (SEQ ID NO. 86), ZBP1-BCAS1 (SEQ ID NO. 95), EBF1-ATP10B (SEQ ID NO. 104), PIGF-CHMP3 (SEQ ID NO. 113), GRLF1-DLGAP1 (SEQ ID NO. 122), E2F3-FLJ22536 (SEQ ID NO. 130), and GPR160-NCEH1 (SEQ ID NO. 139).

The inventors interestingly found that these fusion genes are well suited markers for the diagnosis of lung cancer.

2. Material and Methods

2.1 Break-apart FISH (ba-FISH)

FISH (fluorescence *in situ* hybridization) is a cytogenetic technique that is used to detect and localize the presence or absence of specific DNA sequences on chromosomes. FISH uses fluorescent probes that bind to only those parts of the chromosome with which they show a high degree of sequence similarity. Fluorescence microscopy can be used to find out where the fluorescent probes are bound to the chromosomes.

Break-apart FISH is used for detecting a split between two genes, i.e. two probes that in the wild-type arrangement bind to neighboring positions of a gene, to distant positions. Differently-colored probes were used. Colors that are adjacent appear to overlap; a secondary color is observed. So if two differently-colored probes (break-apart probes) bind at positions that are in close proximity the secondary color will be present. If they bind at remote positions only the primary colors will be present. Locus-specific probes are made for one side of the breakpoint and the other intact chromosome. In normal cells, the secondary colour is observed, but only the primary colour is observed when the translocation occurs. Two color break-apart probes are designed for the sequences of the known breaking point of the chromosome so that a gene translocation can be seen as two separate colors while normal genes are detected as the mixture of two colors. Theoretically, gene translocation FISH assays with break-apart probe sets make it easier to recognize chromosome aberrations since the two colors are seen separately.

To further confirm chimeric transcripts at a genomic level, two unique break-apart FISH assays were designed for each fusion-candidate. For each gene, two bacterial artificial chromosome (BAC) FISH probes were created to hybridize with the neighboring centromeric and telomeric

regions of the gene.

For each gene, the centromeric BAC clone was biotin-14-deoxycytidine triphosphate (dCTP)-labeled (red signal), and the telomeric BAC clone was digoxigenin-dUTP-labeled (green signal). A summary of the BAC clones is displayed below:

5

GENE	Centromeric BAC clone	Telomeric BAC clone
SOS1	RP11-708G9	CTD-2026H4
ADCY3	CTD-3252C16	RP11-1109B14
ZNF142	CTD-3251J21	CTD-232803
PTK6	RP11-939M14	CTD-3228E10
POMC	RP11-159D24	CTD-2359022
E2F4	RP11-620P11	RP11-121M19
EML4	RP11-368J11	RP11-142M12

All of the BACs displayed were derived from the RPCI-11 BAC library (RP11-No.) or from the Caltech-D BAC library (CTD-No.)

The in situ hybridization was performed by methods known in the art: Deparaffinized sections were pretreated with a 100 mM Tris and 50 mM EDTA solution at 92.8°C for 15 min and
 10 digested with Digest-All III (dilution, 1:2;) at 37°C for 14 min; FISH probes were denatured at 73°C for 5 min and immediately placed on ice. Subsequently, the tissue sections and FISH probes were co-denatured at 94°C for 3 min and hybridized overnight at 37°C. After hybridization, washing was done with 2× SSC at 75°C for 5 min, and the fluorescence detection was performed with streptavidin–Alexa 594 conjugates (dilution 1:200) and
 15 antibodies to digoxigenin–fluorescein isothiocyanate (FITC) (dilution, 1:200). Slides were then counterstained with 4',6-diamidino-2-phenylindole (DAPI) and mounted. The samples were analyzed under a 63× oil immersion objective with a fluorescence microscope (Zeiss) equipped with appropriate filters, a charge-coupled device camera, and the FISH imaging and capturing software Metafer 4 (Metasystems). The evaluation of the tests was done independently by three
 20 experienced evaluators. At least 100 nuclei per case were evaluated.

In the case of SOS1-ADCY3 a ba-FISH was performed in the H3122 cell-line where the fusion gene was discovered. For this, metaphase spreads were prepared by treating cells with colcemid (Roche, Switzerland) at a concentration of 0.1 µg/ml. The culture flasks were incubated at

37°C for 1 hour. Mitotic arrest was followed by treatment with 0.075M KCl. The cells were then repeatedly washed with modified Carnoy's fixative (3:1 methanol to acetic acid) and then dropped onto pre-cleaned slides and observed for metaphase spreads. Metaphase spreads were pre-treated with 2x SSC solution at 37°C for 60 min and digested with Digest-All III (dilution 1:2) at 37°C for 6 min.; FISH probes were denatured at 73°C for 5 min. and immediately placed on ice. Subsequently, the metaphase spreads and FISH probes were co-denatured at 85°C for 4 min. and hybridized overnight at 37°C. Post hybridization washing was done with 0.5x SSC at 75°C for 5 min., and the fluorescence detection was carried out using streptavidin-Alexa-594 conjugates (dilution 1:200) and anti-digoxigenin-FITC (dilution 1:200). Slides were then counterstained with 4',6-Diamidin-2' phenylindoldihydrochlorid (DAPI) and mounted. The samples were analysed under a 63x oil immersion objective using a fluorescence microscope (Zeiss, Jena, Germany) equipped with appropriate filters, a charge-coupled device camera and the FISH imaging and capturing software Metafer 4 (Metasystems, Altlussheim, Germany). Assessment of the experiments was done independently by two evaluators.

Rearrangement of either of the genes was defined *a priori* as following: A split of a signal pair, resulting in a single red and single green signal for at least one allele is referred to as a translocation. A loss of a signal, resulting in either a single red or single green signal for at least one allele is referred to as a rearrangement through deletion. A wild-type allele displays a juxtaposed red and green signal (forming a yellow signal).

2.2 Cloning, retrovirus production, and transduction

Total RNA was extracted from frozen tumor sections containing more than 90% tumor cells, using Qiagen RNeasy Mini Kit (Qiagen) and 1µg RNA was reverse transcribed using the SuperScript III Reverse Transcriptase kit (Invitrogen), according to the manufacturer's instructions. cDNA was cleaned by using the Nucleo Spin Extract II Kit (Macherey-Nagel) according to the manufacturer's instructions. Chimeric transcripts were amplified by RT-PCR from the cDNA using candidate-specific primers given in Table 1. Resulting chimeric fragments were cloned into pBabe-puromycin plasmid by using restriction enzyme sites.

4µg pBabe vector containing the sequence of interest together with 4µg pCL vector were co-transfected into HEK293T packaging cells (ATCC), using TransIT-TKO transfection agent

(Mirus). After 48 hours, viral supernatants were filtered, mixed with medium containing total concentration of 8µg/ml polybrene (Sigma) and incubated with NIH-3T3 and Ba/F3 cells (both from ATCC) for 10 hours, in separated flasks. After 12 hours, cells were transferred into medium containing 3µg/ml puromycin.

5

Table 1: Primers for amplification of candidate fusion genes

SEQ ID NO.	Primer name	Sequence (5' to 3')
150	SOS_5_amp	CAGGGATCCATGCAGGCGCAGCAG
151	ADCY3_3_amp	CAGGAATTCTCAGGAGTTGTCCACCACCT
152	ZNF142_5_amp	ATGCGTCGACATTCCCCAGAACAGCGATA
153	PTK6_3_amp	CGTAGTCGACTGCACCCATCACCTCAGTA
154	AGAP1_5_amp	ACCATGAACTACCAGCAGCA
155	IGFBP2_3_amp	GCTGGCTGCGGTCTACTG
156	FNDC3B_5_amp	ATGAATGTACGTCACAATGATG
157	TNIK_3_amp	GCCATGAAGATAAGTGCCAAG
158	C12orf11_5_amp	AGGACCAGGCACGAAAGTTA
159	RASSF8_3_amp	TCTGTTGGGTCTCCTCCCTA
160	POMC_5_amp (break-point)	TTCAAAAACGCCATCATCAA
161	ADD1_3_amp (break-point)	GGCAACAAAACGACACAGAA
162	E2F4_5_amp (break-point)	CCTGGGACTGATAGCAAGGA
163	RPL14_3_amp (break-point)	TGACCCTTCTGAGCTTTTGG

2.3 Soft agar assay (test for oncogenicity)

10 The NIH-3T3 cells, an established line of mouse fibroblast cell, are extremely sensitive to "contact inhibition" of cell division. Under the usual culture conditions, NIH-3T3 cells grow rapidly in sparse culture, but cell division stops after the cells become confluent, at approximately 10⁶ cells per 6-cm dish. The cell monolayer has a typical "cobblestone" appearance. Ba/F3, an IL-3 dependent pro B lymphocyte line, has absolute dependence on

15 murine IL-3 for growth. Both phenotypes can be modified by transducing the cells with a

strong oncogene, so NIH-3T3 cells do not stop growing and Ba/F3 cells become independent of IL3.

In order to test the loss of "contact inhibition" of cell division of NIH-3T3 cells, a colony formation assay in soft agar was performed after transduction of the cells with a fusion candidate. The formation of colonies indicates oncogenic properties of the fusion gene.

NIH-3T3 cells containing different constructs (empty pBabe vector as negative control, KRAS or EML4-ALK as positive control and the candidate of interest) were suspended in growth media containing 10% calf serum (CS) and 0.6% agar and plated in triplicate on 50 µl of solidified growth medium (10% CS; 1.0% agar). Colonies were analyzed with the Scanalyzer imaging system (LemnaTec) or counted by hand.

2.4 Immunoblotting

To assess protein levels and/or protein phosphorylation status, cells were plated in growth medium and lysed after 24h of incubation using 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM Na₂EDTA, 1 mM EGTA, 1% Triton X-100 2.5 mM sodium pyrophosphate, 1 mM beta-glycerophosphate, 1 mM Na₃VO₄, 1 µg/ml leupeptin, 1 mM PMSF, complete Protease Inhibitors Cocktail and Phosphatase Inhibitors Cocktail Set III (Merck) Lysates were cleared after 15 minutes of incubation at 4°C by centrifugation (14000 rpm for 20 minutes, 4°C) and boiled with SDS sample buffer (5x SDS sample buffer: 250 mM Tris (pH 6.8), 50% glycerol, 3.575 M β-mercaptoethanol, 10% SDS and 0.05% bromophenolblue) for 10 minutes. Lysates were separated by 4-12% SDS-PAGE, transferred to nitrocellulose membranes and detected by immunoblotting. Western blotting was performed using the following antibodies: anti-pALK, anti-pAKT, anti-pCREB, anti-CREB and anti-pERK (Cell Signaling), anti-Actin (MP Biomedical), anti-total-ALK (Bethyl Laboratories), anti-pPKA and anti-PKA (Santa Cruz), and anti-SOS1 (BD Laboratories).

2.5 Cell-based screening

To assess whether the fusion genes have effects on the susceptibility of the cells to certain compounds the effect of the compounds on the viability of cells expressing such fusion genes was tested in a cell-based screening.

5

All compounds were purchased from commercial suppliers, dissolved in DMSO, and stored at -20°C . Cells were plated into sterile 96-well microtiter plates using a multichannel pipette and cultured overnight as known in the art. Compounds were then added in serial dilutions. Cellular viability was determined after 96 hours of treatment by measuring cellular ATP content using the CellTiter-Glo Assay (Promega) according to the manufacturer's instructions. Plates were measured on a Mithras LB 940 Plate Reader (Berthold Technologies). GI50 values were determined using R package (Frommolt P, Thomas RK, 2008, Standardized high-throughput evaluation of cell-based compound screens. BMC Bioinformatics. 9:475).

15 2.6 Generation of H3122 PR1 cells

H3122 cells were seeded at $\sim 70\%$ confluence in 6-well dishes in RPMI 1640 with 10% FBS. ALK inhibitor PF02341066 was added at a starting concentration of 30 nM, and cells were maintained in fresh drug-containing medium changed every ~ 72 hours. Drug concentrations were increased as soon as the cells reached $\sim 90\%$ confluence. The concentration of PF02341066 was increased until a final concentration of 3 μM . The resulting pool of resistant cells (designated H3122 PR1) were maintained in RPMI with 10% FBS containing 3 μM PF02341066 or 2 μM Crizotinib (active enantiomer of PF02341066).

25 **3. SOS1-ADCY3**

3.1 Oncogenicity

The amplified cDNA of SOS1-ADCY3 (SEQ ID NO. 9) was cloned into pBabe-puromycin plasmid using the restriction enzymes *Bam*HI (5') and *Eco*RI (3'). The resulting Vector (pBABE-puro-SOS1-ADCY1) is depicted in Figure 1.

30

To assess the oncogenicity of SOS1-ADCY3, viruses containing pBABE-puro-SOS1-ADCY1 were produced and used to transduce NIH-3T3 cells as outlined above. These cell were tested in soft agar assay for colony formation as outlined above. The results are shown in Figure 2, right side. As a negative control NIH-3T3 were transduced with a virus containing the empty pBABE-puromycin vector (empty vector, Figure 2, left side). As a positive control NIH-3T3 cells were transduced with a virus containing a pBABE-puromycin vector for expressing KRAS (a known oncogenes) (Figure 2, middle).

Cells expressing the SOS1-ADCY3 fusion gene showed colony formation, demonstrating the oncogenicity of the fusion gene.

3.2 Formation of foci in culture

NIH 3T3 cells transduced with pBabe empty vector of SOS1-ADCY3 were seeded at ~70% confluence on 6cm cell culture dishes and grown for 10 days with medium change after every ~72h. Cells transformed with the empty vector (e.v.) showed confluent growth while cells expressing SOS1-ADCY3 ectopically exhibited the formation of foci, showing the oncogenicity of the fusion gene (Figure 4).

3.3 Tumor formation in mice

To further validate oncogenicity of a SOS1-ADCY3 fusion gene (SEQ ID NO. 9), NIH-3T3 cells transduced with as outlined above. As a negative control NIH-3T3 were transduced with a virus containing the empty pBABE-puromycin vector (empty vector). As a positive control NIH-3T3 cells were transduced with a virus containing a pBABE-puromycin vector expressing an EML4-ALK fusion gene (a known oncogene). The transduced cells were resuspended in plain RPMI in a concentration of 100×10^6 cells/ml. 5×10^6 cells were injected subcutaneously in mice, under 2,5% Isoflurane-anaesthesia. Tumor growth was monitored every 2nd day by calimetric measurement. Tumor volumes were calculated by $V = (\text{long diameter})^2 \times (\text{short diameter})$. The results are shown in Figure 5. The NIH-3T3 cells transduced with the SOS1-ADCY3 carrying vector induced a tumor development in nude mice at the same rate as the well-known transforming fusion gene EML4-ALK, which confirms the high

oncogenicity of SOS1-ADCY3 fusion genes.

3.4 Sensitivity to MEK-Inhibitors

5 To investigate whether the expression of a SOS1-ADCY3 fusion gene (SEQ ID NO. 9) has influence on the susceptibility of cancer cells to MEK-inhibitors, mice carrying tumors as outlined above (see 3.3) where treated with the MEK inhibitor (MEKi) PD0325901 Therapy in mice harbouring the EML4-ALK and SOS-ADCY3 tumors was started when tumors reached a size of 100mm³. They were treated daily by oral gavage either with 25mg/kg PD-0325901
10 dissolved in vehicle (0,5% methylcellulose + 0,2% Tween80) or with the vehicle-detergent alone. After 14 days of therapy mice were sacrificed and tumors were resected and fixed in 4% formaldehyde for histology. Tumor size was measured using the method described in section 3.3, *supra*, 2, 4, 6 and 8 days post injection. Results are shown in Figure 6 and show that tumors induced by SOS1-ADCY3 expression are sensitive to the MEK inhibitor PD0325901
15 and thus susceptible to MEK inhibitors.

3.5 Activated pathways

To further elucidate the pathways involved in oncogenicity of SOS1-ADCY3 fusion gene the
20 phosphorylation of members of cancer related signalling pathways were investigated. Therefore, NIH-3T3 cells transduced with either a pBABE-puromycin vector expressing an EML4-ALK fusion gene or pBABE-puromycin vector expressing the preferred SOS1-ADCY3 fusion gene (SEQ ID NO. 9). Activation status of PKA, CREB, ERK and AKT was tested by immunoblotting using total cell extracts. Results are shown in Figure 7.

25

PKA is a protein kinase activated by cAMP which is a secondary messenger produced by all cyclases including ADCY3. The expression of a SOS1-ADCY3 fusion gene leads to the hyperphosphorylation of PKA compared to the phosphorylation of PKA in EML4-ALK cell lines (Figure 7) while the amount of total PKA is comparable (Figure 7). Phosphorylated PKA
30 (pPKA) can hence be used as a predictive marker for SOS1-ADCY3 positive tumors. Furthermore, these results show that SOS1-ADCY3 induced tumors are targets of PKA inhibitors.

NIH-3T3 cells carrying SOS1-ADCY3 also showed a strong signal for pCREB which is a direct target of pPKA (Figure 7). CREB-binding protein (CREBBP) binds pCREB and is required for its full activity. Histone Acetyl Transferase (HAT) inhibitors are known specific inhibitors for CREBBP as well as its analogue, EP300. Hence, SOS1-ADCY3 can be used as an predictor for susceptibility of cells, cell cultures or tissues for HAT inhibitors (HATi).

pPKA is also known to activate RAF-MEK-ERK pathway, which explains the sensitivity to MEK inhibitors (see 3.4 above) and the strong signal for pERK in NIH-3T3 expressing SOS1-ADCY3 (Figure 7).

The SOS1-ADCY3 fusion gene is oncogenic by itself, so it can be an independent driver event, the fact that SOS1-ADCY3 can be found in EML4-ALK positive tumour suggested that SOS1-ADCY3 could be a resistant mechanism to ALK inhibitors. In order to test this hypothesis resistant clone (PR1) have been generated by treating the parental ALK-inhibitor-sensitive H3122 cell-line with increasing concentrations of the ALK inhibitor TAE684, for a period of 6 months. Resistance to TAE684 of H3122-PR1 (PR1) and the parental H3122 cell-line was tested at different concentrations from 0.02 μ M to 2 μ M (Figure 8a). H3122-PR1 (PR1) a higher resistance to the ALK inhibitor TAE684 than the parental cell-line (Figure 8a). H3122 and H3122-PR1 were then tested for protein expression of SOS1-ADCY3 (using anti-SOS1) and activation of different targets by immunoblotting (Figure 8b). Interestingly, PKA and ERK were activated in PR1 but not in the parental H3122 (Figure 8b). This is in line with the fact that SOS1-ADCY3 protein expression is higher in the resistant clone (Figure 8b). However, RNA-seq data showed that SOS1-ADCY3 gene was equally expressed in both PR1- and parental-H3122 cell-lines (coverage=24), suggesting the involvement of post-translational modifications in this change on expression.

In summary, the presence of SOS1-ADCY3 allows the identification of tumours that are likely to respond to PKA inhibitors, HAT inhibitors and MEK inhibitors or, to a combination of these ones.

3.6 break-apart FISH

SOS1-ADCY3 was initially detected in the NSCLC cell-line H3122. The translocation was confirmed by break-apart *in situ* hybridisation (Ba-FISH) analysis. Subsequently, 226 adenocarcinoma lung tumors were screened also using ba-FISH analysis. In two tumor samples
5 an aberrant ba-FISH signal was detected for ADCY3 suggesting a recurrent event (Figure 9). These samples did not show a rearrangement of the SOS1 gene indicating that these fusions have different 5' partners.

3.6 Correlation with the EML4-ALK rearrangement

10 It was found by the inventors that the genes involved in EML4-ALK and SOS1-ADCY3 fusions were located in the same region of chromosome 2. To confirm the hypothesis of the inventors that SOS1-ADCY3 is a consequence of EML4-ALK rearrangement. To this end break-apart FISH (ba-FISH) ALK and ADCY3 separately and in combination, in metaphase
15 chromosomes from the cell-line H3122 were performed as outlined above. ALK ba-FISH generated one normal signal, and one single green signal with loss of the red signal. The same pattern was observed when performing ADCY3 ba-FISH. When performing both assays together, the same pattern was observed, i.e. only one green signals were observed per allele (see Figure 16 C). If the two rearrangements occurred in different alleles, two separate green
20 signals on different chromosomes, one for ALK and one for ADCY3, would have been expected.

To determine the functional impact of this chimeric transcript on EML4-ALK positive samples, NIH-3T3 cells carrying EML4-ALK were transduced with SOS1-ADCY3 (pBABE-puro-
25 SOS1-ADCY3). As a control NIH-3T3 cells only carrying EML4-ALK were used. The expression of the fusion gene (SEQ ID NO. 1) was detected using RT-PCR with the primers as disclose herein above. The result is shown in Figure 17. While the signal for the SOS1-ADCY3 fusion gene was detectable two days after transduction in the SOS1-ADCY3 transduced cells it was lost after two weeks. That shows that cells containing only EML4-ALK overgrew EML4-
30 ALK and SOS1-ADCY3 containing. This shows that SOS1-ADCY3 influences the oncogenic activity of EML4-ALK.

4. ZNF142-PTK6

The inventors detected in large-cell neuroendocrine tumor and in lung adenocarcinoma (AD) primary tumour the presence of a novel type of somatic fusion polynucleotide comprising a fragment of ZNF142 fused with a fragment of PTK6 (herein referred to as ZNF142-PTK6), through the next generation sequencing technique RNA-seq. This fusion gene is generated by a genomic rearrangement between chromosome 2 and chromosome 20 and the rearrangement has been validated by RT-PCR. The predicted PROSITE domains of the fusion are shown in Figure 10.

The preferred fusion protein ZNF142-PTK6 (SEQ ID NO. 20) comprises a new kinase domain instead of the original PTK6 kinase domain. While the new kinase domain still comprises the active site of the PTK6 kinase domain the SH2 domain located upstream of the kinase was deleted and exchanged by a sequence comprising three Zink finger domains. The sequence of the new kinase domain is (SEQ ID NO 38).

A ZNF142-PTK6 fusion gene was cloned into pBABE/puro plasmid SuperScript III Reverse Transcriptase First-Strand Synthesis System (Invitrogen) was used for producing ssDNA following the protocol. ZNF-PTK6 was amplified with the Primers ZNF142_5_amp (5'-ATGCGTCGACATTCCCCAGAACAGCGATA-3'; SEQ ID NO. 152), and PTK6_3_amp (5'-CGTAGTCGACTGCACCCATCACCTCAGTA-3'; SEQ ID NO. 153) using the following PCR protocol; 92°C 2min., 30x 92°C 2min. 56°C 15sec 68°C 5min. 30sec.. PCR was run several times in parallel and than products were purified using Nucleo Spin Extract II Kit (Machery-Nagel) according to the manufacturer's instructions. Resulting ZNF-PTK6 DNA fragments were treated with Klenow Fragment (3'->5' exo-) using following conditions: 32µL DNA, 5µL Klenow Buffer, 1mM dATP and 3µL Klenow Exo- (all from NEB) 30min at 37°C. Product was cleaned again by using the Nucleo Spin Extract II Kit (Macherey-Nagel) according to the manufacturer's instructions. Resulting ZNF-PTK6 fragments were cloned into Topo cloning vector using the TA-Cloning-Kit (Invitrogen), sequenced and subcloned into pBabe puro by using the restriction enzyme *SaII*. 4µg pBabe ZNF-PTK6 vector together with 4µg pCL vector were co-transfected into Hek293T packaging cells (ATCC), using TransIT-TKO transfection agent (Mirus). After 48 hours, viral supernatants were filtered, mixed with medium

containing total concentration of 8µg/ml polybrene (Sigma) and incubated with NIH3T3 and Baf3 cells (both from ATCC) for 10 hours in separated flasks. After 12 hours, cells were transferred into medium containing 3µg/ml puromycin. NIH-3T3 transduced with the pBABE/puro plasmid comprising ZNF142-PTK6 fusion gene (SEQ ID NO. 19) cells showed colony formation in soft agar validating the oncogenicity of the fusion gene (Figure 11a). The cells were further tested for tumor induction in mice. To this end the cells and a control were injected into nude mice and tumor formation was monitored over a period of 24 days using as outlined for SOS1-ADCY3 mice experiments above. The results show that mice injected with the NIH-3T3 cells containing the pBABE/puro plasmid comprising ZNF142-PTK6 fusion gene (SEQ ID NO. 19) exhibited detectable tumor growth 16 days post injection while the control mice did not show tumors (Figure 11b).

A set of 373 primary lung tumours, both SQ and AD, was screened using ba-FISH. The identification of 2 out of 160 in SQ (1.2%) rearrangements involving the PTK6 region, and 10 out of 213 AD (4.4%) demonstrates recurrence. Three different patterns were observed (Figure 12): translocation (left panel), deletion of one signal plus amplification of the other one (central panel) and, deletion of one signal (right panel) without amplification.

Furthermore, the samples were tested for rearrangement involving the ZNF142 region. Recurrence could be demonstrated by finding rearrangements in two addition samples (1 SQ (Figure 13, left panel) and 1 AD (Figure 13, right panel) both with deletion of one signal). The rearrangement in AD was in the same sample in which PTK6 translocation was detected (*supra*; Figure 12, left panel).

Ba/F3, an IL-3 dependent pro B lymphocyte line, has absolute dependence on murine IL-3 for growth. This phenotype can be modified by transducing the cells with a strong oncogene, so these cells become independent of IL3. To test for the oncogenic properties of the ZNF142-PTK6 fusion gene, Ba/F3 cells were transduced with an expression vector carrying the ZNF142-PTK6 (SEQ ID NO. 19) fusion gene. The transduced cell, unlike the cells transduced with the empty vector showed growth independent of IL-3 showing the oncogenic properties of the fusion gene.

4. POMC-ADD1

The inventors identified an oncogenic fusion gene comprising the coding sequence of POMC (SEQ ID NO. 33) and the 3'UTR of ADD1 (SEQ ID NO. 37). Ba-FISH assay for POMC in the specific sample showed that this chimeric transcript was not the result of a genomic event; however, since the fusion gene was validated by RT-PCR, this suggests that it could be the result of a trans-splicing event. Levels of POMC gene expression in the POMC-ADD1 tumor samples were compared with their matched-normals and with the levels on non-carriers tumor samples, using the primers summarized in the following table:

SEQ ID NO.	Primer name	Sequence
164	POMC_f	CTGGCCTTGCTGCTTCAG
165	POMC_r	GAAGTGGCCCATGACGTACT

The fusion of POMC coding sequence and 3'UTR of ADD1 significantly increased the expression of POMC when comparing to the matched-normal (Figure 14, left panel), and also when comparing to 5 fusion-negative CAs where POMC was undetectable by RT-PCR (Figure 14, right panel).

5. FNDC3B-TNIK

The inventors detected in a lung AD primary tumour, and through the next generation sequencing technique RNA-seq, the presence of a novel fusion polynucleotide of a partial FNDC3B fused with a partial TNIK (hereinafter referred to as FNDC3B-TNIK) (SEQ ID NO. 4), which is generated by an intra-chromosomal rearrangement in chromosome 3. The rearrangement was confirmed by RT-PCR. The inventors successfully isolated the cDNA of the FNDC3B-TNIK fusion gene and the whole ORF of FNDC3B-TNIK was sequenced. The fusion cDNA was cloned into pBABE/puro plasmid using TopoCloning approach. Viruses containing the plasmid were produced and used to transduce NIH-3T3 and Ba/F3 cells using common methods known by ordinary person in the field. FNDC3B-TNIK-transduced NIH-3T3 cells are able to form colonies in soft agar assays (Figure 15). Screening using ba-FISH as well as mice experiments are ongoing. Afterwards, high throughput compound screening and ba-FISH screening in additional primary tumours will be performed focusing on cJun inhibitors since the CNH domain of TNIK (present in the fusion gene) is known to be able to activate c-Jun.

Hence, c-Jun inhibitors are a predictor for the positive response to c-Jun inhibitors.

Figure legend

5

Figure 1: The predicted PROSITE domains of the SOS1-ADCY3 fusion-protein (SEQ ID NO. 10). The sequence of the fusion gene was validated by reverse-Transcriptase PCR (RT-PCR) and the translation was analysed using ExPASy – PROSITE (*prosite.expasy.org*). The sequence has two predicted Guanylate cyclase domains (rhombus) and one DH2 domain (rounded rectangle).

10

Figure 2: Schematic overview of pBABE-puro-SOS1-ADCY3

Figure 3: Soft agar assay using NIH-3T3 cell lines transduced with an empty pBABE-puromycin vector (empty vector), a pBABE-puromycin vector carrying the KRAS-G12V gene under control of an active promoter as a positive control, or pBABE-puromycin-SOS1-ADCY3 (SOS1-ADCY3). If the expressed gene or fusion gene is oncogenic, the cells are able to form colonies in soft agar.

15

Figure 4: NIH-3T3 cell line transduced with pBABE-puromycin-SOS1-ADCY3 (SOS1-ADCY3) shows formation of foci in culture, while NIH-3T3 cell line transduced with the empty pBAB-puromycin vector (e.v.) shows confluent growth.

20

Figure 5: Tumor induction in mice. NIH-3T3 cells were transduced with either the empty pBABE-puromycin vector (NIH-3T3_e.V., diamonds), with pBABE-puromycin vector carrying the known oncogenic EML4-ALK fusion gene (NIH-3T3_EML4ALK_v.1, circles) or with pBABE-puromycin-SOS1-ADCY3 (NIH-3T3_SOS1-ADCY3; triangles). The cells were then injected into naked mice. Tumor size was measured as described before in each cell line. The averages are shown. Bars indicate standard deviation.

25

Figure 6: Sensitivity of SOS1-ADCY3 induced tumors to MEKi. NIH-3T3 cells were transduced with either pBABE-puromycin vector carrying the known oncogenic EML4-ALK

30

fusion gene (EML4ALK; diamonds) or with pBABE-puromycin-SOS1-ADCY3 (SOS1-ADCY3; triangles). The cell lines were injected into nude mice. These mice were then treated with PD0325901 (PD-treated; dark diamonds or triangles) or control (light diamonds or triangles) as described above. The averages are shown. Bars indicate standard deviation.

5

Figure 7: SOS1-ADCY3 induced phosphorylation of protein kinase A (PKA), cAMP response element-binding protein (CREB), Proteinkinase B (AKT) and extracellular-signal-regulated kinase (ERK) shown by immunoblotting of the respective proteins

10 **Figure 8:** SOS1-ADCY3 induces resistance of tumor cells to ALK inhibitors. H3122-PR1 (PR1) shows a higher resistance to the ALK inhibitor TAE684 than the parental cell-line H3122 (A). PKA and ERK were activated in PR1 but not in the parental H3122 (B). This is in line with the fact that SOS1-ADCY3 protein expression is higher in the resistant clone (B).

15 **Figure 9:** Picture of one of the two ba-FISH ADCY3-positives samples. The arrow indicates the loss of red signal showing that this gene is involved in a translocation in this sample.

Figure 10: The predicted PROSITE domains of the preferred ZNF142-PTK6 fusion protein (SEQ ID NO. 20). The sequence of the fusion gene was validated by reverse-Transcriptase PCR (RT-PCR) and the translation was analysed using ExPASy – PROSITE (prosite.expasy.org). The sequence has 15 predicted Zink finger domains (pentagons) and one novel protein kinase domain (elipse).

20

Figure 11: Oncogenic activity of the ZNF142-PTK6 fusion gene (SEQ ID NO.19). (A) soft agar assay using NIH-3T3 cell lines transduced with an empty pBABE-puromycin vector (empty vector), a pBABE-puromycin vector carrying the KRAS-G12V gene under control of an active promoter as a positive control, or pBABE-puromycin-ZNF142-PTK6 (comprising the ZNF142-PTK6 fusion gene (SEQ ID NO. 19). If the expressed gene or fusion gene is oncogenic, the cells are able to form colonies in soft agar. The number of colonies were counted for each condition and normalized with the positive control. (B) Tumor induction in mice. NIH-3T3 cells were transduced with either the empty pBABE-puromycin vector (NIH-3T3_e.V., diamonds) or with pBABE-puromycin-ZNF142-PTK6 (NIH-3T3_ZNF142-

25

30

PTK6_old, rectangles; NIH-3T3_ZNF142-new; circles) (old and new refers to two different populations of NIH-3T3 transduced with the same construct, replicates of the experiment). The cells were then injected into nude mice. Tumor size was measured as described before.

5 **Figure 12:** Pictures of three ba-FISH PTK6-positive samples. The arrow indicates the patterns of aberrant signals: loss of the red signal (left panel), amplification of the green with loss of the red (middle panel) and green and red break-apart signals (right panel). These aberrant signal suggest that PTK6 is translocated in these samples.

10 **Figure 13:** picture of the two ba-FISH ZNF142-positive samples. The arrows indicate the loss of a signal suggesting that this gene is involved in a translocation in these two samples.

Figure 14: POMC expression in POMC-ADD1 positive tumors (T) carriers versus matched-normal (N) (left panel) and in POMC-ADD1-positive tumors (S01746 and S01502) versus
15 POMC-ADD1-negative tumors (S01501, S01504, S01666, S01731, S01733 and S01733) (right panel).

Figure 15: Soft agar assay using NIH-3T3 cell lines transduced with an empty pBABE-puromycin vector (empty vector), a pBABE-puromycin vector carrying the EML4-ALK fusion
20 gene under control of an active promoter as a positive control, or pBABE-puromycin-FNDC3B-TNIK (carrying the FNDC3B-TNIK fusion gene (SEQ ID NO. 48). If the expressed gene or fusion gene is oncogenic, the cells are able to form colonies in soft agar.

Figure 16: Pictures of one of the ba-FISH of Example 3.6: (A) ADCY3 break-apart assay
25 (Probes 1 and 2); the arrow indicates a chromosome/allele where the red signal is lost and only a green signal could be detected (indicating an rearrangement); (B) ALK break-apart assay (Probes 3 and 4); the arrow indicates a chromosome/allele where the red signal is lost and only a green signal could be detected (indicating an rearrangement); (C) ALK + ADCY3 break apart assay (Probes 1, 2, 3 and 4); the arrow indicates a chromosome/allele where the red signal is
30 lost and only a green signal could be detected (indicating an rearrangement for both; ALK and ADCY3).

Figure 17: RT-PCR gel picture for detection of SOS1-ADCY3 fusion gene expression in pBABE-puro-SOS1-ADCY3 transduced NIH-3T3 cells containing EML3-ALK fusion genes. Cells were grown in puromycin containing media.

Table 2: Sequences disclosed in this application: Sequence identifications referred herein comprise the following sequences and are also identified in the appended sequence protocol

SEQ ID NO.	Name and further information
1	SOS1 (CCDS nucleotide sequence of SOS1 (Gene ID: 6654))
2	SOS1 (full-length protein)
3	SOS1 (preferred gene fragment)
4	SOS 1 (preferred protein fragment)
5	ADCY3 (CCDS nucleotide sequence of ADCY3 (Gene ID: 109))
6	ADCY3 (full-length protein)
7	ADCY3 (preferred gene fragment)
8	ADCY3 (preferred protein fragment)
9	SOS1-ADCY3 (preferred fusion gene)
10	SOS1-ADCY3 (preferred fusion protein)
11	ZNF142 (CCDS nucleotide sequence of ZNF142 (Gene ID: 7701))
12	ZNF142 (full-length protein)
13	ZNF142 (preferred gene fragment)
14	ZNF142 (preferred protein fragment)
15	PTK6 (CCDS nucleotide sequence of PTK6 (Gene ID: 5753))
16	PTK6 (full-length protein)
17	PTK6 (preferred gene fragment)
18	PTK6 (preferred protein fragment)
19	ZNF142-PTK6 (preferred fusion gene)
20	ZNF142-PTK6 (preferred fusion-protein)
38	Novel Kinase Domain
21	AGAP1 (nucleotide sequence of AGAP1 Gene ID: 116987)
22	AGAP1 (full-length protein)
23	AGAP1 (preferred gene fragment)
24	AGAP1 (preferred protein fragment)

25	IGFBP2 (CCDS nucleotide sequence of IGFBP2 (Gene ID: 3485))
26	IGFBP2 (full-length protein)
27	IGFBP2 (preferred gene fragment)
28	IGFBP2 (preferred protein fragment)
29	AGAP1- IGFBP2 (preferred fusion gene)
30	AGAP1- IGFBP2 (preferred fusion-protein)
31	POMC (CCDS nucleotide sequence of POMC (Gene ID: 5443) plus 3'UTR)
32	POMC (full-length protein)
33	POMC (preferred gene fragment)
34	POMC (preferred protein fragment)
35	ADD1 (CCDS nucleotide sequence of ADD1 (Gene ID: 118))
36	ADD1 (full-length protein)
37	ADD1 (preferred gene fragment)
39	POMC-ADD1 (preferred fusion gene)
40	FNDC3B (CCDS nucleotide sequence of FNDC3B (Gene ID: 64778))
41	FNDC3B (full-length protein)
42	FNDC3B (preferred gene fragment)
43	FNDC3B (preferred protein fragment)
44	TNIK (CCDS nucleotide sequence of TNIK (Gene ID: 23043))
45	TNIK (full-length protein)
46	TNIK (preferred gene fragment)
47	TNIK (preferred protein fragment)
48	FNDC3B-TNIK (preferred fusion gene)
49	FNDC3B-TNIK (preferred fusion protein)
50	C12orf11 (CCDS nucleotide sequence of C12orf11 (Gene ID: 55726))
51	C12orf11 (full-length protein)
52	C12orf11 (preferred gene fragment)
53	C12orf11 (preferred protein fragment)

54	RASSF8 (CCDS nucleotide sequence of RASSF8 (Gene ID: 11228))
55	RASSF8 (full-length protein)
56	RASSF8 (preferred gene fragment)
57	RASSF8 (preferred protein fragment)
58	C12orf11-RASSF8 (preferred fusion gene)
59	C12orf11-RASSF8 (preferred fusion protein)
60	E2F4 (CCDS nucleotide sequence of E2F4 (Gene ID: 1874))
61	E2F4 (full-length protein)
62	E2F4 (preferred gene fragment)
63	E2F4 (preferred protein fragment)
64	RPL14 (CCDS nucleotide sequence of RPL14 (Gene ID: 9045))
65	RPL14 (full-length protein)
66	RPL14 (preferred gene fragment)
67	RPL14 (preferred protein fragment)
68	E2F4-RPL14 (preferred fusion gene)
69	E2F4-RPL14 (preferred fusion protein)
70	CHEK2 (CCDS nucleotide sequence of CHEK2 (Gene ID: 11200))
71	CHEK2 (full-length protein)
72	CHEK2 (preferred gene fragment)
73	CHEK2 (preferred protein fragment)
74	THOC5 (CCDS nucleotide sequence of THOC5 (Gene ID: 8563))
75	THOC5 (full-length protein)
76	THOC5 (preferred gene fragment)
77	CHEK2-THOC5 (preferred fusion gene)
78	CHEK2-THOC5 (preferred fusion protein)
79	M6PR (CCDS nucleotide sequence of M6PR (Gene ID: 4074))
80	M6PR (full-length protein)
81	M6PR (preferred gene fragment)
82	M6PR (preferred protein fragment)

83	CD4 (CCDS nucleotide sequence of CD4 (Gene ID: 920))
84	CD4 (full-length protein)
85	CD4 (preferred gene fragment)
86	M6PR-CD4 (preferred fusion gene)
87	M6PR-CD4 (preferred fusion protein)
88	ZBP1 (CCDS nucleotide sequence of ZBP1 (Gene ID: 81030))
89	ZBP1 (full-length protein)
90	ZBP1 (preferred gene fragment)
91	ZBP1 (preferred protein fragment)
92	BCAS1 (CCDS nucleotide sequence of BCAS1 (Gene ID: 8537))
93	BCAS1 (full-length protein)
94	BCAS1 (preferred gene fragment)
95	ZBP1-BCAS1 (preferred fusion gene)
96	ZBP1-BCAS1 (preferred fusion protein)
97	EBF1 (CCDS nucleotide sequence of EBF1 (Gene ID: 1879))
98	EBF1 (full-length protein)
99	EBF1 (preferred gene fragment)
100	EBF1 (preferred protein fragment)
101	ATP10B (CCDS nucleotide sequence of ATP10B (Gene ID: 23120) plus 395 nucleotides of the 5'UTR)
102	ATP10B (full-length protein)
103	ATP10B (preferred gene fragment)
104	EBF1-ATP10B (preferred fusion gene)
105	EBF1-ATP10B (preferred fusion protein)
106	PIGF (CCDS nucleotide sequence of PIGF (Gene ID: 5281) including 150 nucleotides of the 5'UTR)
107	PIGF (full-length protein)
108	PIGF (preferred gene fragment (part of the 5'UTR))
110	CHMP3 (CHMP3 (Gene ID: 51652))
111	CHMP3 (full-length protein)
112	CHMP3 (preferred gene fragment)

113	PIGF-CHMP3 (preferred fusion gene)
115	GRLF1 (GRLF1 (Gene ID: 2909))
116	GRLF1 (full-length protein)
117	GRLF1 (preferred gene fragment)
118	GRLF1 (preferred protein fragment)
119	DLGAP1 (CCDS nucleotide sequence of DLGAP1 (Gene ID: 9229))
120	DLGAP1 (full-length protein)
121	DLGAP1 (preferred gene fragment)
122	GRLF1-DLGAP1 (preferred fusion gene)
123	GRLF1-DLGAP1 (preferred fusion protein)
124	E2F3 (CCDS nucleotide sequence of E2F3 (Gene ID: 1871))
125	E2F3 (full-length protein)
126	E2F3 (preferred gene fragment)
127	E2F3 (preferred protein fragment)
128	FLJ22536 (long intergenic non-protein coding RNA 340; CCDS nucleotide sequence of FLJ22536 (Gene ID: 401237))
129	FLJ22536 (preferred fragment)
130	E2F3-FLJ22536 (preferred fusion gene)
131	E2F3-FLJ22536 (preferred fusion protein)
132	GPR160 (CCDS nucleotide sequence of GPR160 (Gene ID: 26996) including 416 nucleotides of the 5'UTR)
133	GPR160 (full-length protein)
134	GPR160 (preferred fragment (part of the 5'UTR))
136	NCEH1 (CCDS nucleotide sequence of NCEH1 (Gene ID: 57552))
137	NCEH1 (full-length protein)
138	NCEH1 (preferred gene fragment)
139	GPR160-NCEH1 (preferred fusion gene)
141	ADCY1 (full length protein of ADCY1 (Gene ID: 107))
142	ADCY2 (full length protein of ADCY2 (Gene ID: 108))

143	ADCY4 (full length protein of ADCY4 (Gene ID: 196883))
144	ADCY5 (full length protein of ADCY5 (Gene ID: 111))
145	ADCY6 (full length protein of ADCY6 (Gene ID: 112))
146	ADCY7 (full length protein of ADCY7 (Gene ID: 113))
147	ADCY8 (full length protein of ADCY8 (Gene ID: 114))
148	ADCY9 (full length protein of ADCY9 (Gene ID: 115))
149	ADCY10 (full length protein of ADCY10 (Gene ID: 55811))

CLAIMS

5

1. A nucleic acid encoding a sequence at least 70% identical to a protein sequence selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 10), ZNF142-PTK6 (SEQ ID NO. 20), and FNDC3B-TNIK (SEQ ID NO. 49).

10

2. A nucleic acid comprising a sequence at least 70% identical to a sequence selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 9), ZNF142-PTK6 (SEQ ID NO. 19), and FNDC3B-TNIK (SEQ ID NO. 48).

15

3. A protein having a sequence with at least 70 % identity to a sequence selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 10), ZNF142-PTK6 (SEQ ID NO. 20), and FNDC3B-TNIK (SEQ ID NO. 49).

20

4. A method for diagnosing lung cancer in a patient comprising the steps of:
 - determining the presence and/or expression of at least one fusion gene in a sample derived from said patient; and
 - attributing the presence or expression of said fusion gene to the presence of lung cancer in said patient;

25

wherein the fusion gene comprises one or more nucleotide sequences encoding a protein or a fragment of the protein, said protein being selected from the group consisting of a guanine nucleotide exchange factor for RAS proteins, an adenylyl cyclase, a Zink finger protein, a protein tyrosine kinase, a fibronectin type III domain containing protein, and a germinal center kinase (GCKs).

30

5. The method according to claim 4, wherein said guanine nucleotide exchange factor for RAS proteins is SOS1 (SEQ ID NO. 2) or a protein having at least 70% identity thereto; wherein said adenylyl cyclase is selected from the group of adenylyl cyclases consisting of ADCY1 (SEQ ID NO. 141), ADCY2 (SEQ ID NO. 142), ADCY3 (SEQ ID NO. 6),

ADCY4 (SEQ ID NO. 143), ADCY5 (SEQ ID NO. 144), ADCY6 (SEQ ID NO. 145), ADCY7 (SEQ ID NO. 146), ADCY8 (SEQ ID NO. 147), ADCY9 (SEQ ID NO. 148), and ADCY10 (SEQ ID NO. 149); or a protein having at least 70% identity thereto;

wherein said Zink finger protein is ZNF142 (SEQ ID NO. 12) or a protein having at least 70% identity thereto;

wherein said protein tyrosine kinase is PTK6 (SEQ ID NO. 16) or a protein having at least 70% identity thereto;

wherein said fibronectin type III domain containing protein is FNDC3B (SEQ ID NO. 41) or a protein having at least 70% identity thereto; and

wherein said germinal center kinase (GCK) is TNIK (SEQ ID NO. 45) or a protein having at least 70% identity thereto.

6. The method according to claim 4 or 5, wherein said fusion gene comprises one or more nucleotide sequence of a gene, or a fragment of the gene, said gene having a sequence with at least 70% identity to a sequence selected from the group consisting of SOS1 (SEQ ID NO. 1), ADCY3 (SEQ ID NO. 5), ZNF142 (SEQ ID NO. 11), PTK6 (SEQ ID NO. 15), FNDC3B (SEQ ID NO. 40), and TNIK (SEQ ID NO. 44).

7. The method according to claim 6, wherein said fusion gene comprises one or more nucleotide sequences of a fragment of the gene, said fragment having a sequence with at least 70% identity to a sequence selected from the group consisting of SOS1 (SEQ ID NO. 3), ADCY3 (SEQ ID NO. 7), ZNF142 (SEQ ID NO. 13), PTK6 (SEQ ID NO. 17), FNDC3B (SEQ ID NO. 42), and TNIK (SEQ ID NO. 46).

8. The method according to any one of claims 1 to 4, wherein the fusion gene is selected from the group of fusion genes consisting of SOS1-ADCY3, ZNF142-PTK6, and FNDC3B-TNIK.

9. The method according to claim 8, wherein the fusion gene has a sequence with at least 70% identity to a sequence selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 9), ZNF142-PTK6 (SEQ ID NO. 19), and FNDC3B-TNIK (SEQ ID NO. 48).

10. The method according to any one of claims 4 to 9, wherein the lung cancer is selected from a group consisting of any of the lung cancer tumor types as classified in the WHO classification of lung tumors (2004), preferably selected from the group consisting of squamous, adenocarcinoma, small cell carcinoma, large cell carcinoma (including large cell neuroendocrine carcinoma), and carcinoid tumor.
11. A method according to any one of claims 4 to 10, wherein the presence and/or expression of said fusion gene is detected by a hybridisation and/or an amplification assay.
12. The method according to claim 11, wherein the hybridisation and/or amplification assay is selected from the group consisting of polymerase chain reaction (PCR), real-time PCR, reverse transcriptase PCR (RT-PCR), fluorescent *in situ* hybridisation (FISH), chromogenic *in situ* hybridisation (CISH), break-apart *in situ* hybridisation (ba-FISH), and silver *in situ* hybridisation (SISH).
13. A method of selecting a cell, tissue or cell culture with susceptibility to a MEK inhibitor, comprising the steps:
- determining the presence and/or expression of a fusion gene in said cell, tissue or cell culture, wherein the fusion gene comprises a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase in said cell, tissue or cell culture; and
 - attributing the presence and/or expression of said fusion gene to the susceptibility of the cell, tissue or cell culture to a MEK inhibitor.
14. A method for identifying a responder or non-responder to a MEK inhibitor, the method comprising the steps:
- providing a sample from a patient suspected to suffer or being prone to suffer from lung cancer; and
 - determining the presence and/or expression of a fusion gene comprising a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase, or a fragment thereof;

wherein the presence and/or expression of said fusion gene is indicative for a responder to a MEK inhibitor.

15. A method of determining resistance of a cell, tissue or cell culture to an ALK inhibitor,
5 comprising the steps:

- determining the presence and/or expression level of a fusion gene in said cell, tissue or cell culture, wherein the fusion gene comprises a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase in said cell, tissue or cell culture;
- 10 - comparing the determined expression level to the expression level in a control cell, tissue or cell culture expressing said fusion gene and not being resistant to an ALK inhibitor; and
- attributing a determined expression level higher than in the control cell, tissue or cell culture to the resistance of the cell, tissue or cell culture to a ALK inhibitor;
- 15 and/or
- attributing a determined expression equal or lower than in the control cell, tissue or cell culture to the susceptibility of the cell, tissue or cell culture to a ALK inhibitor.

- 20 16. A method for identifying a responder or non-responder to an ALK inhibitor, the method comprising the steps:

- determining the presence/expression level of a fusion gene in a sample from a patient suspected to suffer or being prone to suffer from lung cancer, said fusion gene comprising a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase, or a fragment thereof; and
- 25 - comparing the determined expression level to the expression level in a sample from a control patient, wherein the control patient is a responder to ALK inhibitors and expresses said fusion gene;

wherein a determined expression higher than in said sample of a control patient is
30 indicative for a non-responder to a ALK inhibitor; and

wherein a determined expression equal or lower than in said sample of a control patient is indicative for a responder to a ALK inhibitor.

17. A method of selecting a cell, tissue or cell culture with susceptibility to a HAT inhibitor, comprising the steps:

- determining the presence and/or expression of a fusion gene in said cell, tissue or cell culture, wherein the fusion gene comprises a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase in said cell, tissue or cell culture; and
- attributing the presence and/or expression of said fusion gene to the susceptibility of the cell, tissue or cell culture to a HAT inhibitor.

18. A method for identifying a responder or non-responder to a HAT inhibitor, the method comprising the steps:

- determining the presence and/or expression of a fusion gene in a sample from a patient suspected to suffer or being prone to suffer from lung cancer, said fusion gene comprising a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase, or a fragment thereof;

wherein the presence and/or expression of said fusion gene is indicative for a responder to a HAT inhibitor.

19. A method of selecting a cell, tissue or cell culture with susceptibility to a PKA inhibitor, comprising the steps:

- determining the presence and/or expression of a fusion gene in said cell, tissue or cell culture, wherein the fusion gene comprises a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase in said cell, tissue or cell culture; and
- attributing the presence and/or expression of said fusion gene to the susceptibility of the cell, tissue or cell culture to a PKA inhibitor.

20. A method for identifying a responder or non-responder to a PKA inhibitor, the method comprising the steps:

- determining the presence and/or expression of a fusion gene in a sample from a patient suspected to suffer or being prone to suffer from lung cancer, said fusion

gene comprising a nucleotide sequence encoding a protein or a fragment of the protein, said protein being an adenylyl cyclase, or a fragment thereof;
wherein the presence and/or expression of said fusion gene is indicative for a responder to a PKA inhibitor.

- 5
21. A method according to any one of claims 13 to 20, wherein said adenylyl cyclase is selected from the group consisting of ADCY1 (SEQ ID NO. 141), ADCY2 (SEQ ID NO. 142), ADCY3 (SEQ ID NO. 6), ADCY4 (SEQ ID NO. 143), ADCY5 (SEQ ID NO. 144), ADCY6 (SEQ ID NO. 145), ADCY7 (SEQ ID NO. 146), ADCY8 (SEQ ID NO. 147), ADCY9 (SEQ ID NO. 148), and ADCY10 (SEQ ID NO. 149); or a protein having at least 70% identity thereto.
- 10
22. A method according to any one of claims 13 to 21, wherein said fusion gene further comprises a nucleotide sequence encoding a second protein or a fragment of the second protein, said second protein being a guanine nucleotide exchange factor for RAS proteins.
- 15
23. A method according to claim 22, wherein said guanine nucleotide exchange factor for RAS proteins is SOS1 (SEQ ID NO. 2) or a protein having at least 70% identity thereto.
- 20
24. A method according to any one of claims 13 to 23, wherein the fusion gene comprises one or more nucleotide sequence of a gene or a fragment of the gene, said gene having at least 70% identity to a sequence selected from the group consisting of SOS1 (SEQ ID NO. 1), and ADCY3 (SEQ ID NO. 5), or a fragment thereof.
- 25
25. A method according to claim 24, wherein the fusion gene comprises one or more nucleotide sequence of a fragment of the gene, said fragment of the gene being selected from the group consisting of SOS1 (SEQ ID NO. 3), and ADCY3 (SEQ ID NO. 7).
- 30
26. The method according to any one of claims 13 to 25, wherein the fusion gene is a SOS1-ADCY3 fusion gene.

27. The method according to any one of claims 13 to 26, wherein the fusion gene has a sequence at least 70% identical to SOS1-ADCY3 (SEQ ID NO. 9).

28. A method of selecting a cell, tissue or cell culture with susceptibility to a PTK inhibitor, comprising the steps:

- determining the presence and/or expression of a fusion gene in said cell, tissue or cell culture, wherein the fusion gene comprises a nucleotide sequence encoding a protein or a fragment of the protein, said protein being a protein tyrosine kinase, in said cell, tissue or cell culture; and
- attributing the presence and/or expression of said fusion gene to the susceptibility of the cell, tissue or cell culture to a PTK inhibitor.

29. A method for identifying a responder or non-responder to a PTK inhibitor, the method comprising the steps:

- determining the presence and/or expression of a fusion gene in a sample from a patient suspected to suffer or being prone to suffer from lung cancer, said fusion gene comprising a nucleotide sequence encoding a protein or a fragment of the protein, said protein being a protein tyrosine kinase, or a fragment thereof;

wherein the presence and/or expression of said fusion gene is indicative for a responder to a PTK inhibitor.

30. The method according to claim 28 or 29, wherein said protein tyrosine kinase is PTK6 (SEQ ID NO. 16) or a protein having at least 70% identity thereto; and the PTK inhibitor is a PTK6 inhibitor.

31. The method according to any one of claims 28 to 30, wherein the fusion gene further comprises nucleotide sequence encoding a protein or a fragment of the protein, said protein being a Zinc finger protein.

32. The method according to claim 31, wherein said Zinc finger protein is ZNF142 (SEQ ID NO. 12) or a protein having at least 70% identity thereto.

33. The method according to any one of claims 32 to 36, wherein the fusion gene encodes a protein sequence of ZNF142-PTK6 (SEQ ID NO. 20) or a protein having at least 70% identity thereto.
- 5 34. A method of selecting a cell, tissue or cell culture with susceptibility to a c-Jun inhibitor, comprising the steps:
- determining the presence and/or expression of a fusion gene in said cell, tissue or cell culture, wherein the fusion gene comprises a nucleotide sequence encoding a protein or a fragment of the protein, said protein being a germinal center kinase, in said cell, tissue or cell culture; and
 - attributing the presence and/or expression of said fusion gene to the susceptibility of the cell, tissue or cell culture to a c-Jun inhibitor.
- 10
35. A method for identifying a responder or non-responder to a c-Jun inhibitor, the method comprising the steps:
- determining the presence and/or expression of a fusion gene in a sample from a patient suspected to suffer or being prone to suffer from lung cancer, said fusion gene comprising a nucleotide sequence encoding a protein or a fragment of the protein, said protein being germinal center kinase, or a fragment thereof;
- 15
- 20 wherein the presence and/or expression of said fusion gene is indicative for a responder to a c-Jun inhibitor.
36. The method according to claim 34 or 35, wherein said germinal center kinase is TNIK (SEQ ID NO. 45) or a protein having at least 70% identity thereto.
- 25
37. The method according to any one of claims 34 to 36, wherein the fusion gene further comprises nucleotide sequence encoding a protein or a fragment of the protein, said protein being fibronectin type III domain containing protein.
- 30 38. The method according to claim 37, wherein said fibronectin type III domain containing protein is FNDC3B (SEQ ID NO. 41) or a protein having at least 70% identity thereto.

39. The method according to any one of claims 34 to 38, wherein said fusion gene encodes a protein sequence of FNDC3B-TNIK (SEQ ID NO. 49) or a protein having at least 70% identity thereto.

5 40. A method for identifying compounds for the treatment of lung cancer comprising the steps:

- providing a cell expressing a fusion gene, wherein the fusion gene comprises one or more nucleotide sequence encoding a protein or a fragment of the protein, said protein being selected from the group consisting of a guanine nucleotide exchange factor for RAS proteins, an adenylyl cyclase, a Zink finger protein, a protein tyrosine kinase, a fibronectin type III domain containing protein, and a germinal center kinase (GCKs);
- contacting said cell to a candidate compound; and
- evaluating the viability of said cell contacted with said candidate compound.

15 41. The method according to claim 40, wherein the cell is a cell of a cell line selected from the group consisting of lung cancer cell-lines.

20 42. The method according to claim 40 or 41, wherein said guanine nucleotide exchange factor for RAS proteins is SOS1 (SEQ ID NO. 2) or a protein having at least 70% identity thereto;

wherein said adenylyl cyclase is selected from the group of adenylyl cyclises consisting of ADCY1 (SEQ ID NO. 141), ADCY2 (SEQ ID NO. 142), ADCY3 (SEQ ID NO. 6) ADCY4 (SEQ ID NO. 143), ADCY5 (SEQ ID NO. 144), ADCY6 (SEQ ID NO. 145),
25 ADCY7 (SEQ ID NO. 146), ADCY8 (SEQ ID NO. 147), ADCY9 (SEQ ID NO. 148), and ADCY10 (SEQ ID NO. 149); or a protein having at least 70% identity thereto;

wherein said Zink finger protein is ZNF142 (SEQ ID NO. 12) or a protein having at least 70% identity thereto;

30 wherein said protein tyrosine kinase is PTK6 (SEQ ID NO. 16) or a protein having at least 70% identity thereto;

wherein said fibronectin type III domain containing protein is FNDC3B (SEQ ID NO. 41) or a protein having at least 70% identity thereto;

wherein said germinal center kinase (GCK) is TNIK (SEQ ID NO. 45) or a protein having at least 70% identity thereto.

43. The method according to any one of claims 40 to 42, wherein said fusion gene
5 comprises one or more nucleotide sequence of a gene or a fragment of the gene, said gene having a sequence at least 70% identity to a sequence selected from the group consisting of SOS1 (SEQ ID NO. 1), ADCY3 (SEQ ID NO. 5), ZNF142 (SEQ ID NO. 11), PTK6 (SEQ ID NO. 15), FNDC3B (SEQ ID NO. 40), TNIK (SEQ ID NO. 44).
- 10 44. The method according to claim 43, wherein said fusion gene comprises one or more nucleotide sequence of a fragment of the gene, said fragment having a sequence at least 70% identity to a sequence selected from the group consisting of SOS1 (SEQ ID NO. 3), ADCY3 (SEQ ID NO. 7), ZNF142 (SEQ ID NO. 13), PTK6 (SEQ ID NO. 17), FNDC3B (SEQ ID NO. 42), and TNIK (SEQ ID NO. 46).
- 15 45. The method according to any one of claims 40 to 44, wherein said fusion gene is a fusion gene selected from the group of fusion genes consisting of SOS1-ADCY3, ZNF142-PTK6, and FNDC3B-TNIK.
- 20 46. The method according to any one of claims claim 40 to 45, wherein the fusion gene has a sequence with at least 70% identity to a sequence selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 9), ZNF142-PTK6 (SEQ ID NO. 19), and FNDC3B-TNIK (SEQ ID NO. 48).
- 25 47. The use of a cell line expressing at least one of the fusion genes selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 9), ZNF142-PTK6 (SEQ ID NO. 19), and FNDC3B-TNIK (SEQ ID NO. 48) for screening and/or validation of a medicament for the treatment of lung cancer.
- 30 48. The use according to claim 47, wherein the cell is a transgenic cell.
49. The use according to claim 47 or 48, wherein the lung cancer is selected from the group

of lung cancer types consisting of squamous, adenocarcinoma, small cell carcinoma, large cell carcinoma (including large cell neuroendocrine carcinoma), and carcinoid tumor.

- 5 50. A kit useful for carrying out the method of any one of claims 3 to 31, said kit comprising an oligonucleotide or polynucleotide capable of amplifying and/or specifically detecting a fusion gene selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 9), ZNF142-PTK6 (SEQ ID NO. 19), and FNDC3B-TNIK (SEQ ID NO. 48); and/or said kit comprising an antibody specifically binding to a fusion protein selected from the group consisting of SOS1-ADCY3 (SEQ ID NO. 10), ZNF142-PTK6 (SEQ ID NO. 20), and FNDC3B-TNIK (SEQ ID NO. 49).
- 10
51. A method for identifying the presence and expression of a fusion gene comprising a nucleic acid encoding an adenylyl cyclase in a sample, the method comprising the step of:
- 15
- (i) determining the phosphorylation status of PKA and/or CREB;
 - (ii) comparing the determined phosphorylation status to the phosphorylation status of PKA and/or CREB in a control; and
 - (iii) attributing a higher phosphorylation status of PKA and/or CREB in determined
- 20 step (i) than in the control to the presence and expression of a fusion gene comprising a nucleotide sequence encoding a protein, said protein being an adenylyl cyclase.
52. The method according to claim 51, wherein said adenylyl cyclase is selected from the group consisting of ADCY1 (SEQ ID NO. 141), ADCY2 (SEQ ID NO. 142), ADCY3 (SEQ ID NO. 6) ADCY4 (SEQ ID NO. 143), ADCY5 (SEQ ID NO. 144), ADCY6 (SEQ ID NO. 145), ADCY7 (SEQ ID NO. 146), ADCY8 (SEQ ID NO. 147), ADCY9 (SEQ ID NO. 148), and ADCY10 (SEQ ID NO. 149); or a protein having at least 70% identity thereto.
- 25
53. The method according to claim 51 or 52, wherein a higher phosphorylation status of PKA and/or CREB determined in step (i) than in the control is attributed to the presence and
- 30

expression of a fusion gene further comprising a nucleotide sequence encoding a second protein or a fragment of the second protein, said second protein being a guanine nucleotide exchange factor for RAS proteins.

- 5 54. The method according to claim 53, wherein said second protein being a guanine nucleotide exchange factor for RAS proteins is proteins is SOS1 (SEQ ID NO. 2) or a protein having at least 70% identity thereto.
- 10 55. The method according to claim 53 or 54, wherein the fusion gene is a SOS1-ADCY3 fusion gene, preferably a fusion gene having a sequence at least 70% identical to the sequence of SOS1-ADCY3 (SEQ ID NO. 9).
56. The method according to any one of claims 51 to 55, wherein the phosphorylation status is determined by fluorescent *in situ* hybridisation (FISH).

Figure 1



Figure 2

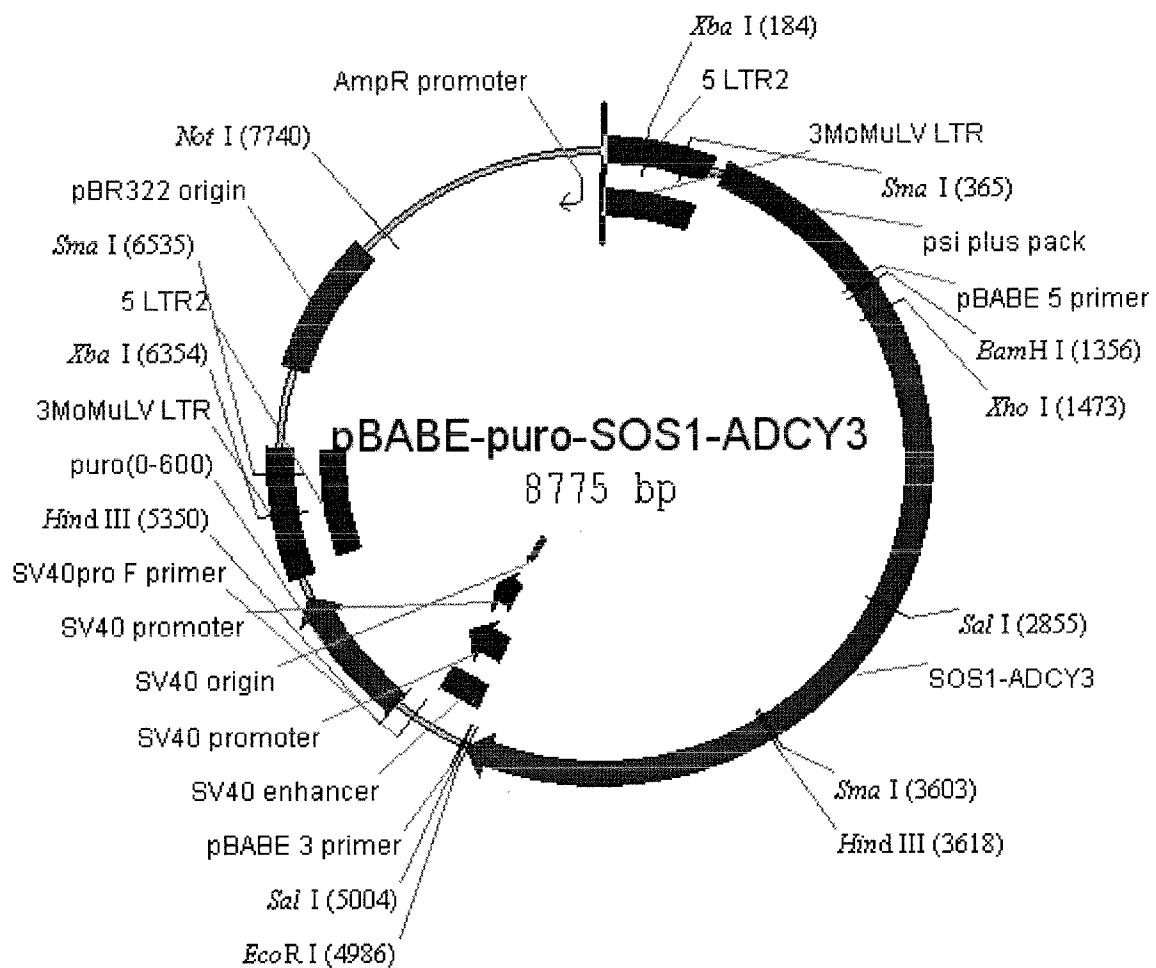
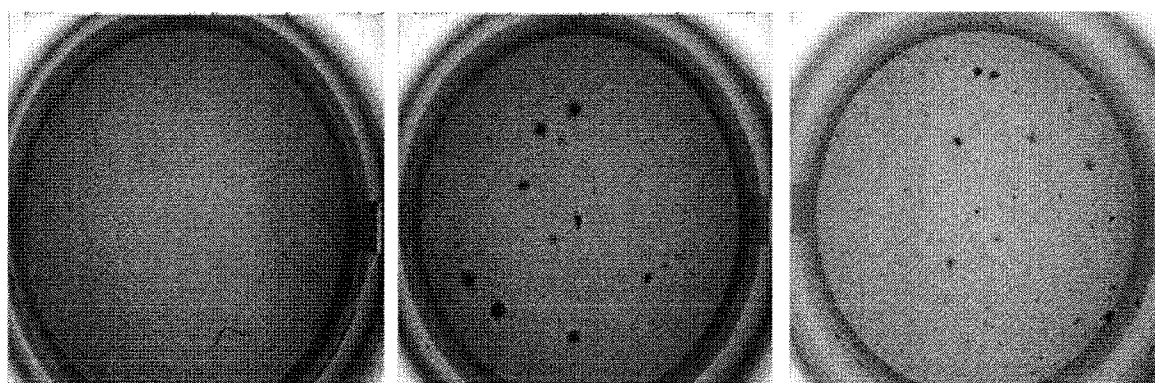
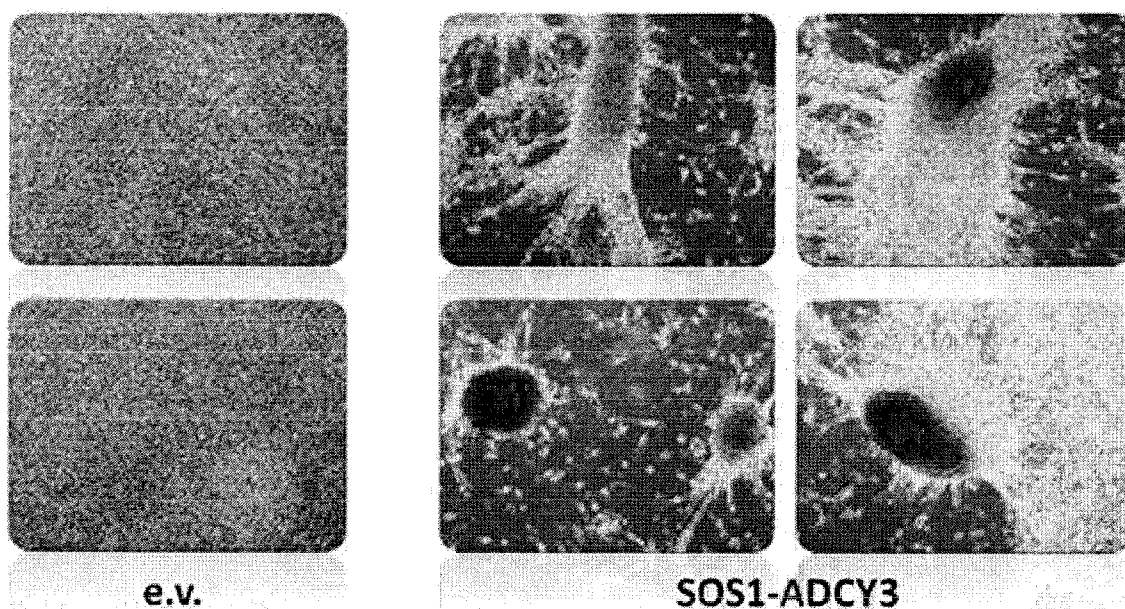


Figure 3

Empty vector

KRAS

SOS1-ADCY3

Figure 4

e.v.

SOS1-ADCY3

Figure 5

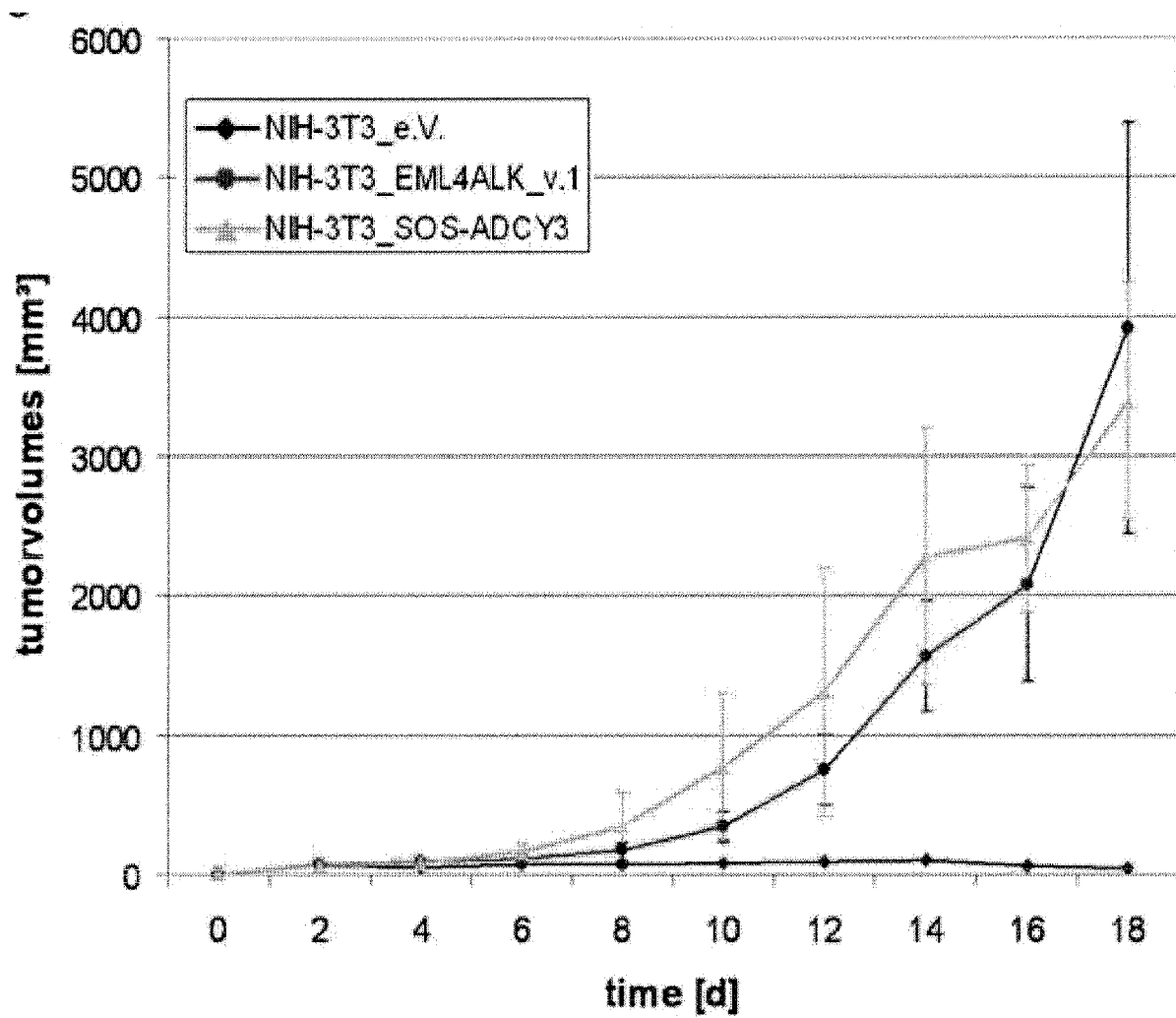


Figure 6:

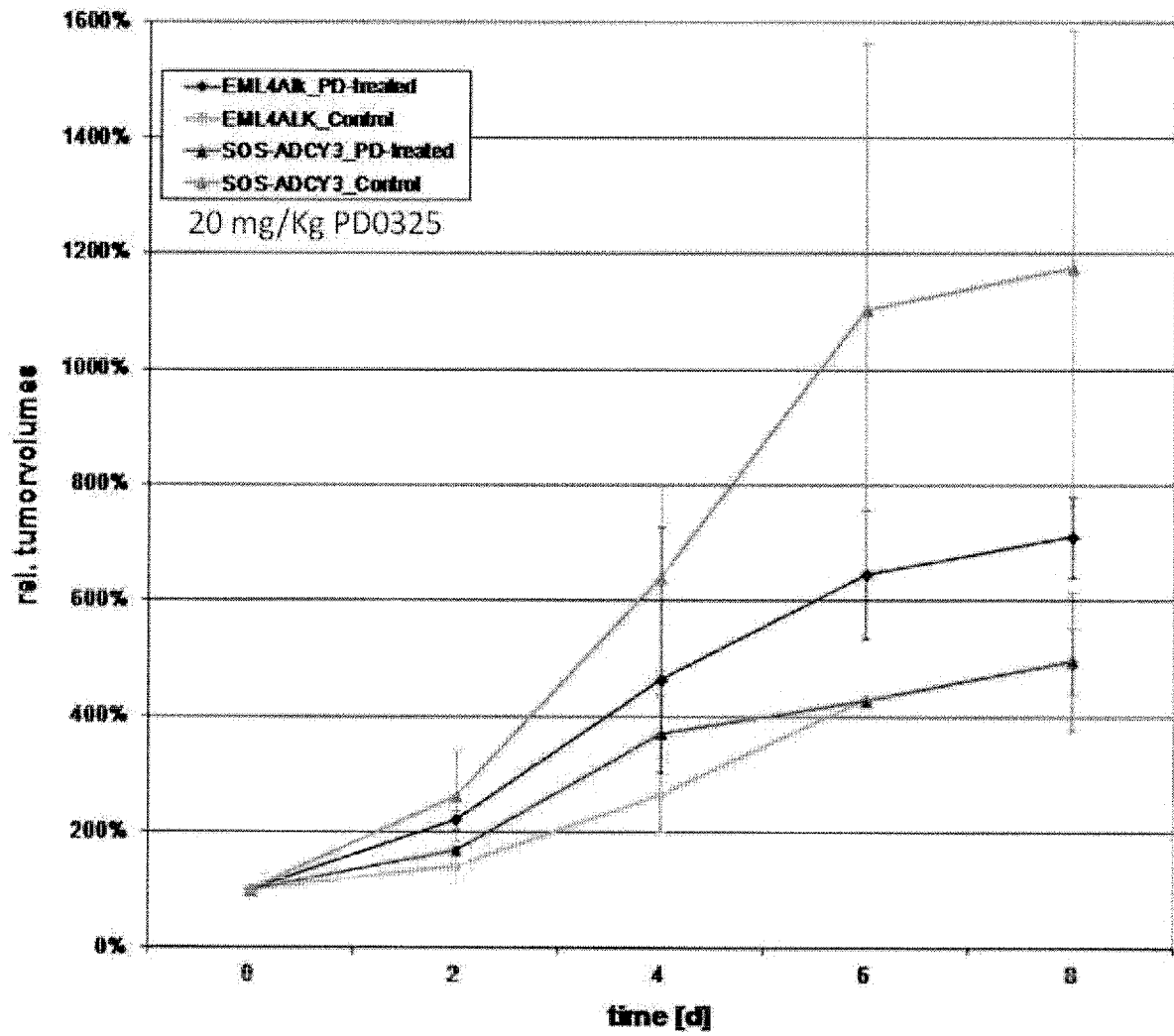


Figure 7

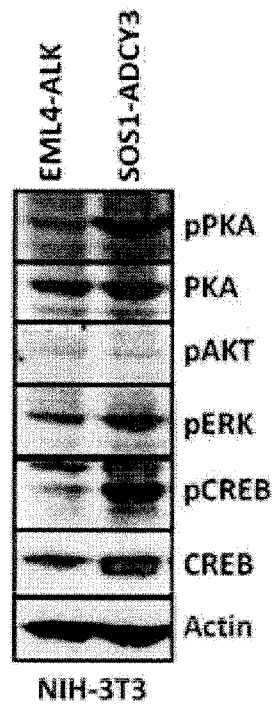


Figure 8:

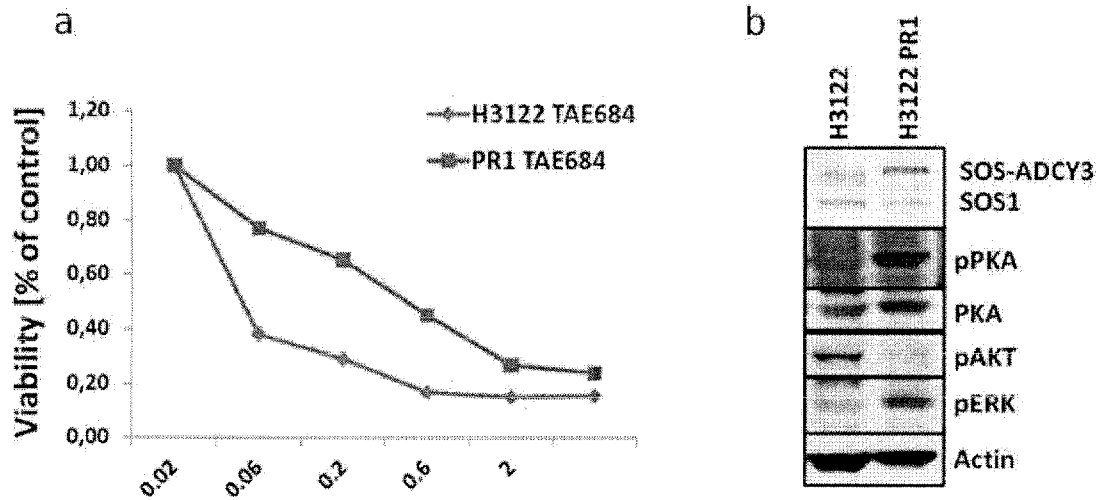
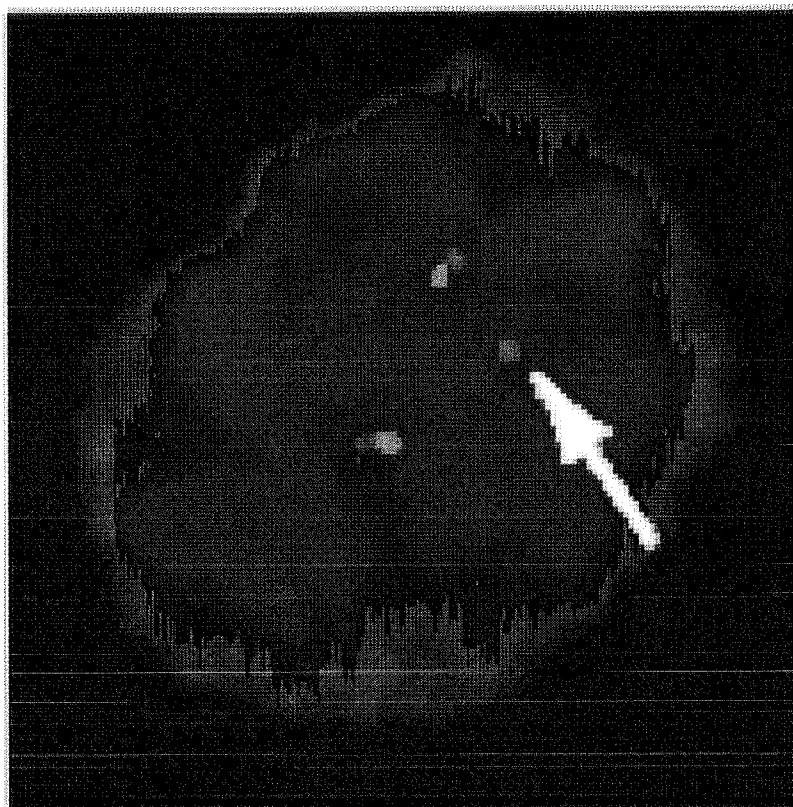


Figure 9



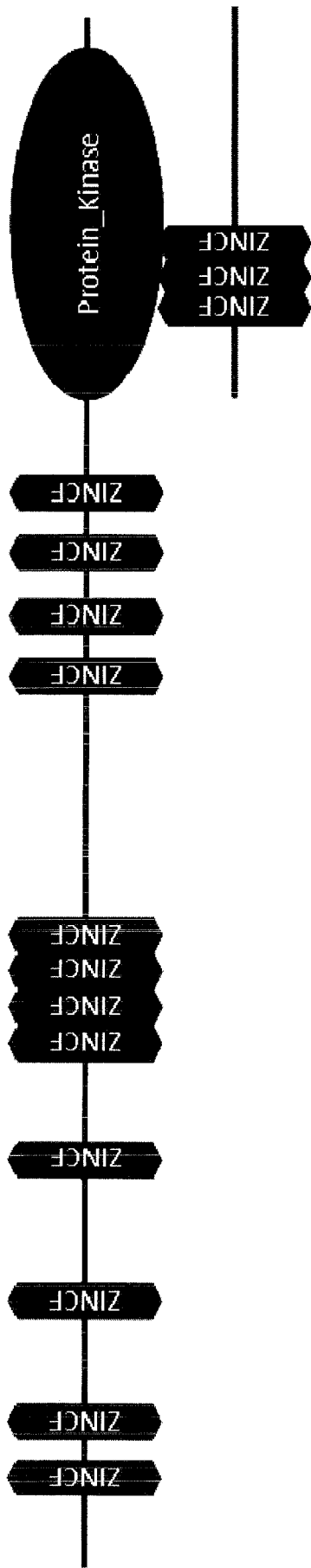
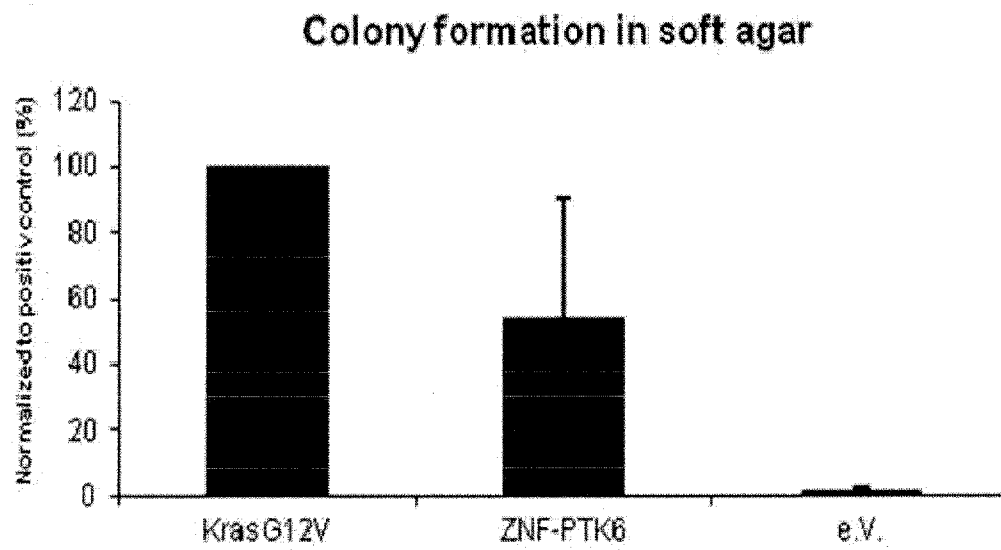


Figure 10

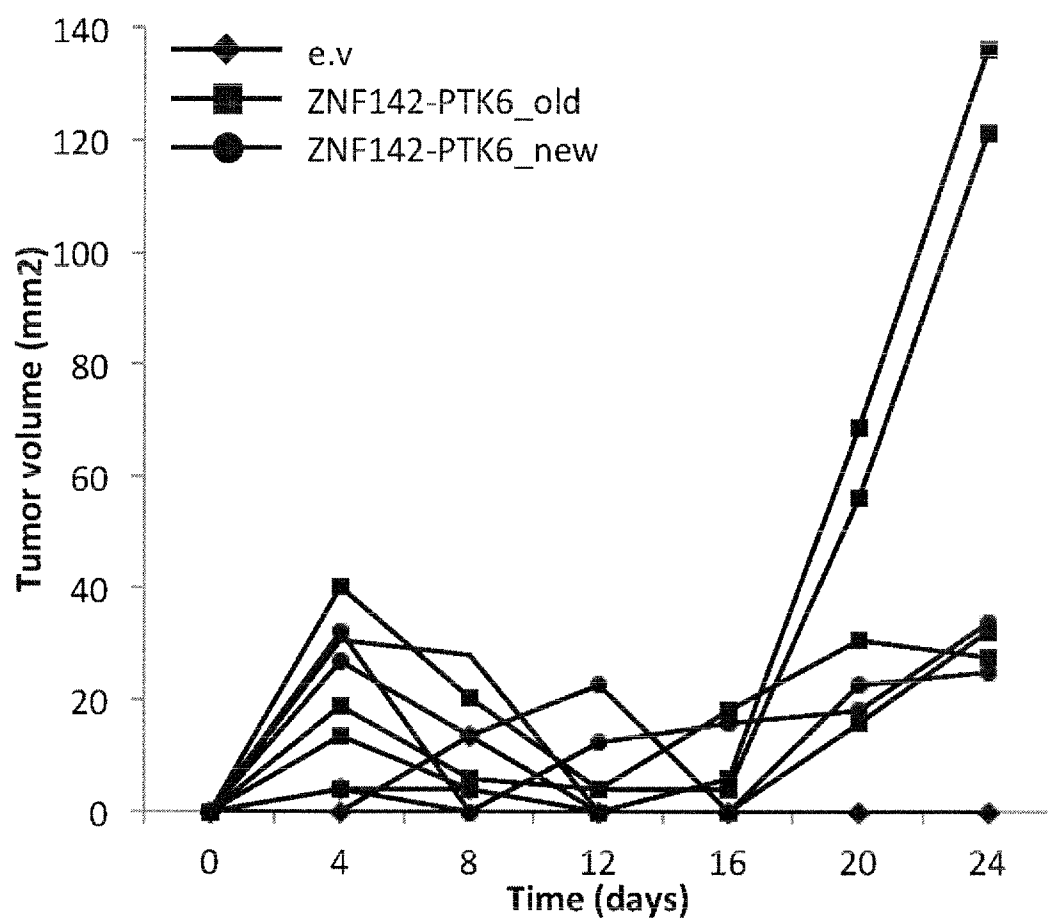
8/12

Figure 11

a



b



9/12

Figure 12

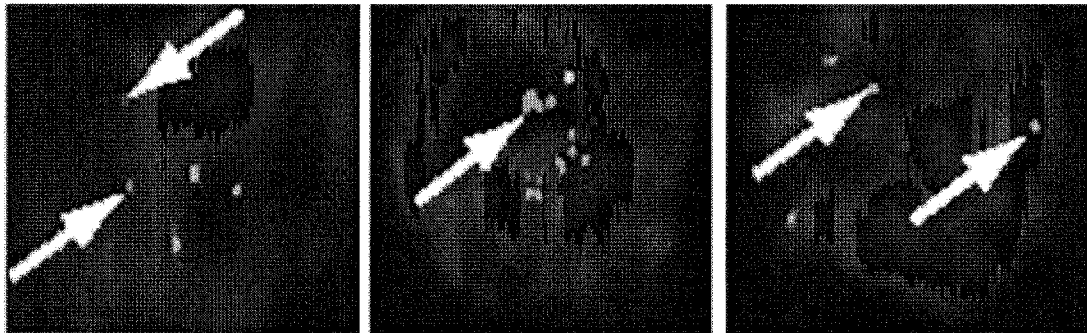


Figure 13

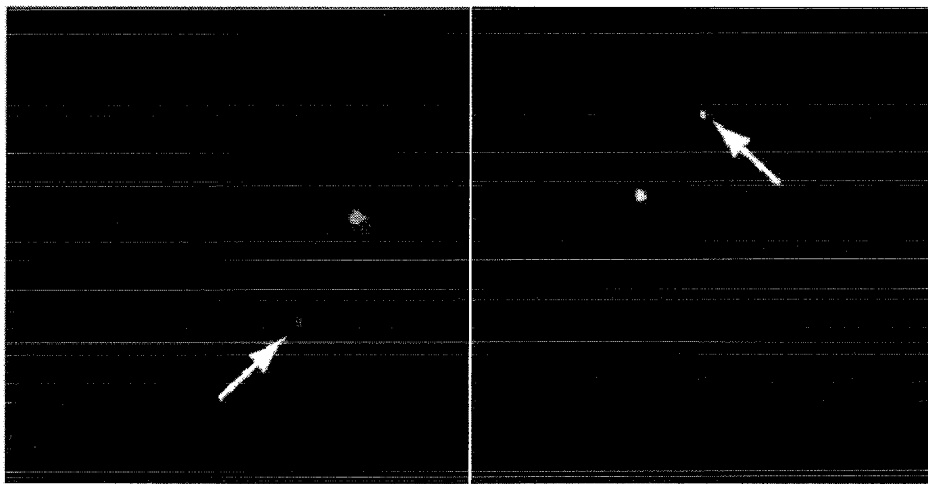


Figure 14

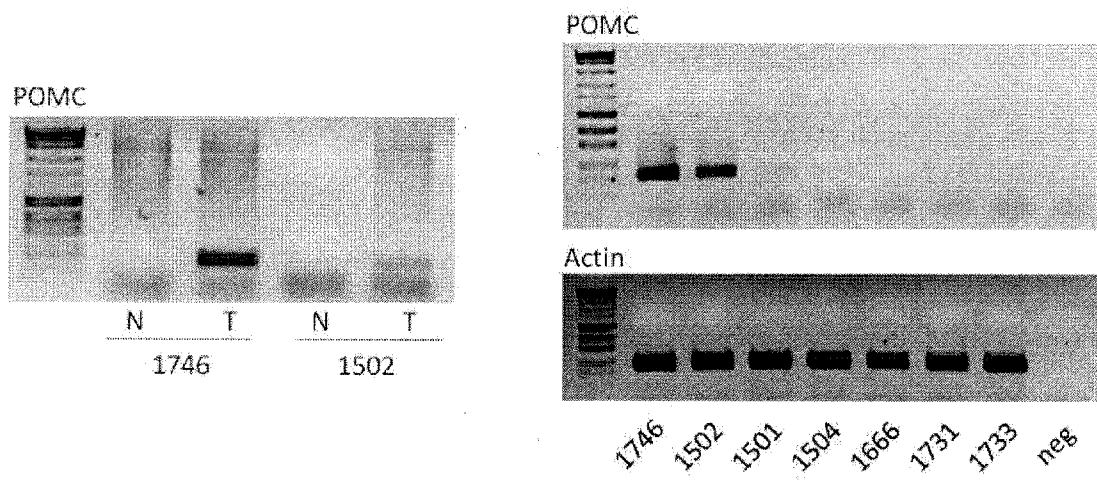
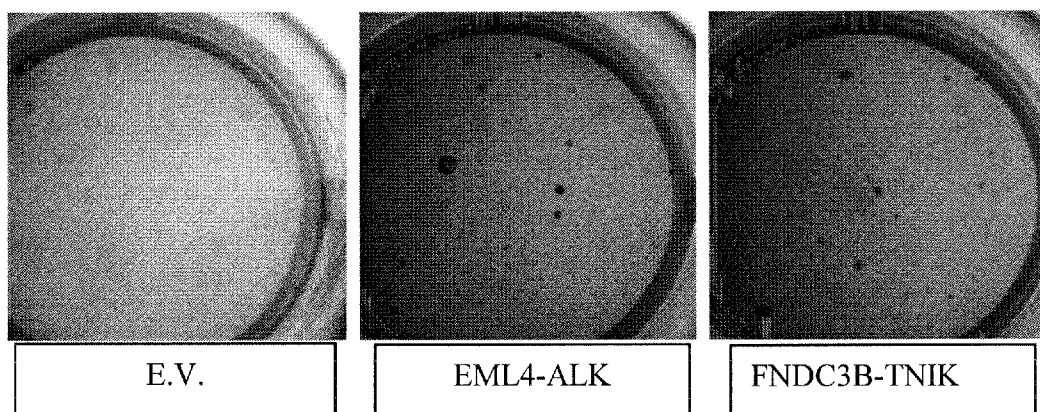


Figure 15



11/12

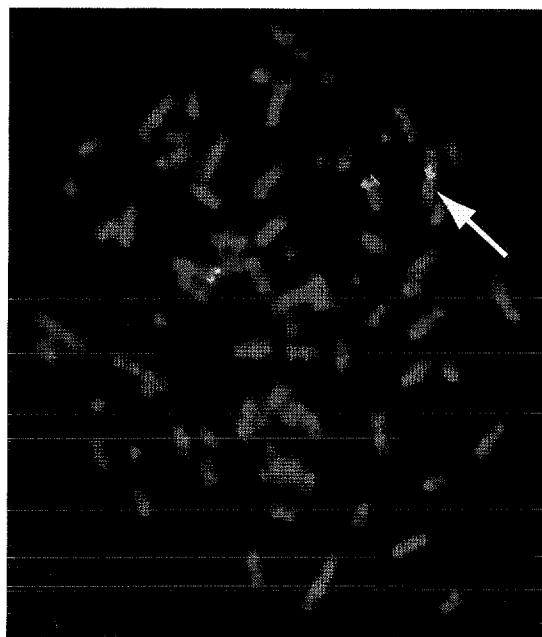
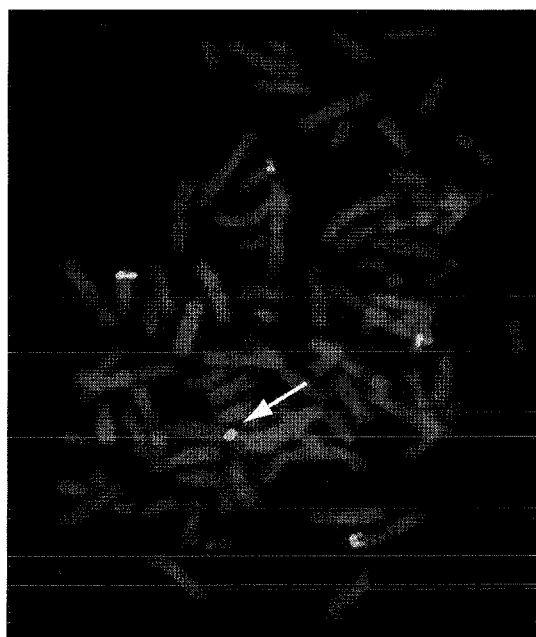
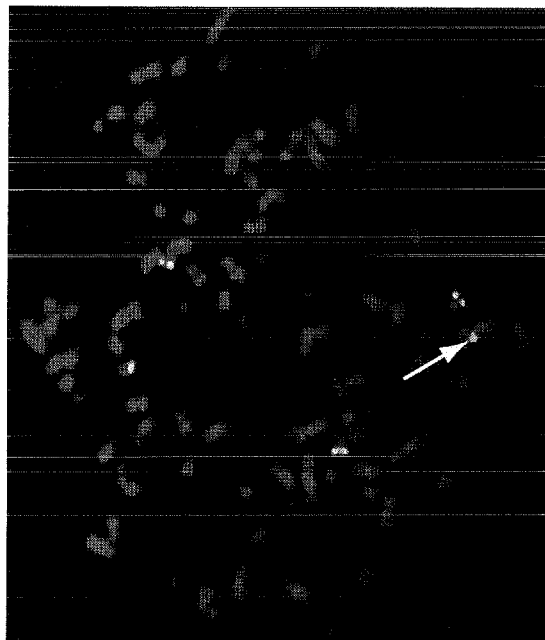
Figure 16**A****B****C**

Figure 17

