FGFR FUSIONS

FIELD OF THE INVENTION

The present invention relates to fusion genes that comprise a polypeptide-coding sequence from a fibroblast growth factor receptor and another polypeptide-coding sequence. It further relates to the fusion polypeptides encoded by the genes as well as to DNA copies of the fusion polypeptide-coding sequences. The invention also encompasses diagnostic and therapeutic applications that are based on the latter fusion genes and polypeptides.

BACKGROUND OF THE INVENTION

Certain somatic fusion genes have been known to be drivers of cancer initiation and progression. Mittelman, F., et al. (2007) Nature Reviews Cancer 7: 233-245. The first, now classic, example of a cancer-promoting fusion gene is the *BCR-ABL1* fusion gene that is found in over 95% of chronic myelogenous leukemia (chronic-phase CML) patients. The *BCR-ABL1* gene encodes a constitutively active form of ABL kinase. The optimal frontline treatment for patients with chronic-phase CML is the subject of active clinical evaluation but involves relatively specific inhibitors of the BCR/ABL tyrosine kinase. Currently marketed inhibitors include first generation drug imatinib (current first line treatment) and second generation drugs nilotinib, dasatinib, bosutinib and ponatinib. Fusion genes were also found to occur with high frequencies in other hematological cancers. Annala, M.J., et al. (2013) Cancer Lett. 340: 192-200. The ETV6-RUNX1 and BCR-ABL1 fusions appear in 25% and 14%, respectively, of acute lymphocytic leukemias, the RUNX1-ETO and CBFB-MYH11 fusions in 10-15% of acute myeloid leukemias, the IG@-MYC fusion in 90-100% of Burkitt's lymphomas, the PML-RARA fusion in 95% of acute promyelocytic leukemias, and the NPM1-ALK and TPM-ALK fusions in 75% and 15%, respectively, of anaplastic large cell lymphomas. While fusion genes historically were detected with relatively high frequencies in hematological cancers, they were only found in a small fraction of solid tumors. More recently, however, it became clear that fusion genes could also occur with elevated frequencies in solid tumors. Annala et al. (2013). Fusions of TMPRSS2 and members of the ETS family

of transcription factors were found in about 70% of prostate cancer patients. EML4-ALK fusions can be present in non-small cell lung cancers, KIAA1549-BRAF fusions in pediatric glioma and FGFR3-TACC3 fusions in glioblastoma. Comprehensive listings of known fusion genes are found, e.g., in Annala et al. or in Shaw, A.T. et al. (2013) Nature Reviews Cancer 13: 772-787. It is noted that some fusions can occur in different cancers. As an example, TPM3-ALK fusions were identified in anaplastic large cell lymphoma and in inflammatory myofibroblastic tumors. Other ALK fusions occur in non-small lung cell cancers as well as in anaplastic large cell lymphoma.

Fusions can be cancer-promoting by different mechanisms. In the case of BCR-ABL, for example, the BCR partner provides dimerization domains, causing constitutive dimerization of the ABL domain, which results in constitutive ABL kinase activity and, consequently, uncontrolled cell division. An alternative mechanism is at play in the case of the TMPRSS2-ETS fusions found in prostate cancer. In these fusion genes, a sequence coding for an ETS transcription factor is brought under the control of the androgen-regulated TMPRSS2 promoter, causing the transcription factor to be overexpressed. Overexpressed ETS dysregulates the expression of genes associated with normal prostate epithelial differentiation and causes uncontrolled cell proliferation. In yet another mechanism, up-regulation of the expression of the FGFR polypeptide can result from the loss of a miRNA regulation site in the 3'UTR of the FGFR mRNA, which loss occurs when the FGFR gene fuses with another gene. Parker, B.C. et al. (2013) J. Clin. Invest. 123: 855–865.

Discovery and characterization of fusion genes advance cancer therapy in multiple ways. Taking as examples fusion genes encoding activated tyrosine kinases, e.g., ABL1, ALK, ROS1, RET and FGFR1-3, identification of such fusion genes in cancerous tissue from patients motivates the discovery and development of selective or specific inhibitors directed against the relevant kinases. The presence of fusion kinase genes also informs the choice of therapeutic approach. For example, the first line treatment for chronic-phase CML patients expressing BCR-ABL1 fusion kinase is a regimen comprising BCR-ABL kinase inhibitor imatinib. Discovery of fusion kinase genes provides a basis for devising diagnostic assays that are capable of discovering the presence of such genes or the expression of the products of such genes in tissues from a cancer patient. As discussed for chronic-phase CML, a positive diagnosis of the presence of a fusion kinase gene or of gene products thereof in a tumor tissue of a patient will allow a

physician to decide on the most appropriate therapy regimen. Typically, such a regimen will include administration of a composition that inhibits the expression or the activity of the fusion kinase in question.

There is no reason to believe that all fusion genes relevant to cancer (or other diseases) are now known. The discovery and characterization of additional fusion genes is expected to increase the specificity of cancer treatment subsequent to the development of diagnostic methods for the newly discovered fusion genes or polypeptides and the development or identification of specific inhibitors of the newly discovered fusion genes or polypeptides or of other agents directed to the fusion genes or polypeptides. In fact, there is an increasing need for identifying specific subpopulations, for example, among cancer patients who would benefit the most from a given treatment such as a therapy involving a particular kinase inhibitor.

SUMMARY OF THE INVENTION

The present invention relates to a cDNA encoding a fusion polypeptide comprising an FGFR2 polypeptide and a CCDC147 polypeptide or a VCL polypeptide, wherein the FGFR2 polypeptide is the whole or a part of a wildtype FGFR2 polypeptide, or the whole or a part of a mutant FGFR2 polypeptide having one or more amino acid substitutions, deletions, or insertions with respect to the wildtype polypeptide; the CCDC147 polypeptide is the whole or a part of a wildtype CCDC147 polypeptide, or the whole or a part of a mutant CCDC147 polypeptide having one or more amino acid substitutions, deletions, or insertions with respect to the wildtype polypeptide; and the VCL polypeptide is the whole or a part of a wildtype VCL polypeptide, or the whole or a part of a mutant VCL polypeptide having one or more amino acid substitutions, deletions, or insertions with respect to the wildtype polypeptide. The FGFR2 polypeptide in a fusion polypeptide can be the whole or a part of a wildtype polypeptide having the amino acid sequence of either of SEQ ID NOs: 17, 25, 27, 29, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57 and 59, or the whole or a part of a mutant polypeptide that differs by one or more amino acid substitutions, deletions, or insertions from the respective wildtype polypeptide which is a polypeptide according to either of SEQ ID NOs: 17, 25, 27, 29, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57 and 59. The CCDC147 polypeptide in a fusion polypeptide can be the whole or a part of a wildtype polypeptide having the amino acid sequence of either of

SEQ ID NOs: 21 and 63, or the whole or a part of a mutant polypeptide with one or more amino acid substitutions, deletions, or insertions in the wildtype polypeptide. The VCL polypeptide in a fusion polypeptide can be the whole or a part of a wildtype polypeptide having the amino acid sequence of either of SEQ ID NOs: 19, 33, 35 and 61, or the whole or a part of a mutant polypeptide with one or more amino acid substitutions, deletions, or insertions in the wildtype polypeptide.

The cDNA encoding a fusion polypeptide comprising an FGFR2 polypeptide and a CCDC147 polypeptide or a VCL polypeptide combines a first and a second polynucleotide. The first polynucleotide encodes the whole or a part of a wildtype FGFR2 polypeptide or a mutant polynucleotide derived therefrom by substitution, deletion or insertion of one or more codons; and the second polynucleotide encodes the whole or a part of a wildtype CCDC147 polypeptide or VCL polypeptide or a mutant polynucleotide derived therefrom by substitution, deletion or insertion of one or more codons. In a particular embodiment, the first polynucleotide encodes the whole or a part of a wildtype FGFR2 polypeptide and includes all or part of the nucleotide sequence of either of SEQ ID NOs: 16, 22, 24, 26, 28, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56 and 58, or a nucleotide sequence derived therefrom by substitution, deletion or insertion of one or more codons. The second polynucleotide encodes the whole or a part of a wildtype CCDC147 polypeptide or a wildtype VCL polypeptide and includes all or part of the nucleotide sequence of either of SEQ ID NOs: 20, 36 and 62 (CCDC147) or either of SEQ ID NOs: 18, 30, 32, 34 and 60 (VCL), or a nucleotide sequence derived from any of the latter nucleotide sequences (i.e., SEQ ID NOs: 20, 36 or 62 (CCDC147), or SEQ ID NOs: 18, 30, 32, 34 or 60 (VCL)) by substitution, deletion or insertion of one or more codons (amino acid-coding nucleotide triplets).

In a more particular embodiment, the cDNA encoding a fusion polypeptide comprises an FGFR2 polypeptide including a complete tyrosine kinase domain and a CCDC147 polypeptide or a VCL polypeptide. A tyrosine kinase domain is considered complete, if it enables the fusion polypeptide to exhibit a detectable tyrosine kinase activity. In this embodiment, the first of the two polynucleotides of the cDNA encoding a fusion polypeptide comprising an FGFR2 polypeptide and a CCDC147 polypeptide or a VCL polypeptide encodes an FGFR2 polypeptide fragment that is sufficiently long to include a complete tyrosine kinase domain.

In an even more particular embodiment, the first of the two polynucleotides of the cDNA encoding a fusion polypeptide comprising an FGFR2 polypeptide and a CCDC147 polypeptide or a VCL polypeptide contains FGFR2 exons 1-16 and part or all of FGFR2 exon 17, and the second of the two polynucleotides contains either CCDC147-coding sequences that lack CCDC147 exon 1 (or that start from exon 2) or VCL-coding sequences that lack VCL exons 1-14 (or that start from exon 15). In the present invention, the exon annotation is performed on the basis of the longest coding transcript found in Ensembl v42 assembly for each part of a fusion gene, i.e. SEQ ID NO: 16 for FGFR2, SEQ ID NO: 18 for VCL and SEQ ID NO: 20 for CCDC147.

The cDNAs of the afore-mentioned embodiments can be derived from a gene transcript isolated from a human cholangiocarcinoma.

More specific embodiments concern the polynucleotide (cDNA) sequences of SEQ ID NO: 1 and SEQ ID NO: 2.

The present invention also encompasses vectors carrying a cDNA of the invention, whereby such vectors can be expression vectors that support the expression of the fusion polypeptide encoded in the cDNA in the cell type for which the vectors are adapted. The invention also relates to any cell (e.g. prokaryotic or eukaryotic) that contains a vector carrying a cDNA of the invention. This cell can be an *E.coli* or a mammalian cell.

Other embodiments of the invention relate to fusion polypeptides comprising an FGFR2 polypeptide and a CCDC147 polypeptide or a VCL polypeptide, which fusion polypeptides are either recombinant polypeptides, polypeptides isolated from cancer cells propagated *in vitro* or as xenografts, or polypeptides isolated (purified) from human cholangiocarcinomas. The FGFR2 polypeptide comprised in the fusion polypeptide is the whole or a part of a wildtype FGFR2 polypeptide, or the whole or a part of a mutant FGFR2 polypeptide having one or more amino acid substitutions, deletions, or insertions with respect to the wildtype polypeptide; the CCDC147 polypeptide is the whole or a part of a wildtype CCDC147 polypeptide, or the whole or a part of a mutant CCDC147 polypeptide having one or more amino acid substitutions, deletions, or insertions with respect to the wildtype polypeptide; and the VCL polypeptide is the whole or a part of a wildtype VCL polypeptide, or a part of a mutant VCL polypeptide having one or more amino acid substitutions, deletions, or insertions with respect to the wildtype polypeptide.

In a particular embodiment, the FGFR2 polypeptide comprised in the fusion polypeptide is the whole or a part of a wildtype polypeptide having the amino acid sequence of either of SEQ ID NOs: 17, 25, 27, 29, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57 and 59, or the whole or a part of a mutant polypeptide having one or more amino acid substitutions, deletions, or insertions with respect to the wildtype polypeptide. The CCDC147 polypeptide comprised in the fusion polypeptide is the whole or a part of a wildtype polypeptide having the amino acid sequence of either of SEQ ID NOs: 21 and 63, or the whole or a part of a mutant polypeptide having one or more amino acid substitutions, deletions, or insertions with respect to the wildtype polypeptide. The VCL polypeptide comprised in the fusion polypeptide is the whole or a part of a wildtype polypeptide having the amino acid sequence of either of SEQ ID NOs: 19, 33, 35 and 61, or the whole or a part of a mutant polypeptide having one or more amino acid substitutions, deletions, or insertions with respect to the wildtype polypeptide.

A further embodiment relates to an antibody or antigen-binding fragment that binds a fusion polypeptide of the invention. The antibody or antigen-binding fragment can recognize any sequence of the fusion polypeptide. In a specific embodiment, the antibody or antigen-binding fragment recognizes an epitope that is composed of sequences from both fusion partners, i.e., from FGFR2 and from CCDC147 or VCL.

The invention also relates to oligonucleotides that can specifically hybridize to a gene for a fusion polypeptide of the invention, an RNA transcript of such a gene or a cDNA of the invention. A particular embodiment concerns a primer pair consisting of a sense and an antisense primer, which primers are capable of specifically hybridizing to a cDNA of the invention and of directing replication of the cDNA or parts thereof. The same primers can also be used to amplify genomic sequences encoding a fusion polypeptide of the invention or RNA transcripts thereof. Another embodiment relates to an oligonucleotide probe capable of specifically hybridizing to a cDNA of the invention, a gene encoding a fusion polypeptide of the invention or an RNA transcript of such a gene. In another embodiment, the oligonucleotide is an antisense oligonucleotide capable of hybridization in a life cell to a messenger RNA encoding a fusion polypeptide of the invention. Such hybridization prevents or reduces translation of the messenger RNA. Alternatively, the oligonucleotide can be a siRNA directed to a messenger RNA encoding a fusion polypeptide of the invention.

The present invention also encompasses kits for detecting either a fusion polypeptide of the invention or a gene encoding such a fusion polypeptide or its RNA transcripts. A kit for detecting a fusion polypeptide can comprise one or more antibodies or antigen-binding fragments capable of binding the fusion polypeptide. A kit for detecting a gene or a transcript of a gene encoding a fusion polypeptide can comprise the above-described primer pair or oligonucleotide probe.

Also within the scope of the present invention are inhibitors of FGFR kinase activity for use in a therapeutic regimen in a subject suffering from cancer, wherein said subject contains or expresses a gene encoding a fusion polypeptide of the invention or expresses such a fusion polypeptide. The FGFR kinase inhibitor can be selected from the group consisting of 5-amino-1-(2-methyl-1H-benzimidazol-5-yl)-1H-pyrazol-4-yl]-(1H-indol-2-yl)-methanone, PD173074, Pazopanib, AZD4547, Ponatinib, Dovitinib, BGJ398, E-3810, JNJ-42756493, ARQ 087, LY2874455, BAY1163877, ASP5878, E7090, ODM-203, Nintedanib, TAS-120, PRN 1109 and PRN 1371. More specifically, the FGFR kinase inhibitor can be 5-amino-1-(2-methyl-1H-benzimidazol-5-yl)-1H-pyrazol-4-yl]-(1H-indol-2-yl)-methanone.

The present invention also concerns a method of personalized cancer therapy, comprising subjecting a subject containing or expressing a gene for a fusion polypeptide of the invention or expressing such a fusion polypeptide to a therapeutic regimen that comprises administration of a pharmaceutical composition comprising (1) an inhibitor of FGFR kinase activity, (2) an antibody or antigen-binding fragment that recognizes the fusion polypeptide, (3) an antisense oligonucleotide capable of hybridization to a messenger RNA encoding a fusion polypeptide of the invention or (4) an siRNA directed to a messenger RNA encoding a fusion polypeptide of the invention.

A method of therapy that also concerns aspects of diagnosis comprises the steps of (a) taking a biopsy or fluid sample containing cancer cells or tumor circulating DNA from a subject suffering from cancer; (b) determining whether the cells in the biopsy or fluid sample contain or express a gene encoding a fusion polypeptide of the invention or express such a fusion polypeptide; (c) selecting the subject containing or expressing the gene for the fusion polypeptide or expressing the fusion polypeptide for the treatment of step d; and (d) subjecting the selected subject to a therapeutic regimen that comprises administration of a pharmaceutical composition comprising (1) an inhibitor of FGFR

kinase activity, (2) an antibody or antigen-binding fragment that recognizes the fusion polypeptide, (3) an antisense oligonucleotide capable of hybridization to a messenger RNA encoding a fusion polypeptide of the invention or (4) an siRNA directed to a messenger RNA encoding a fusion polypeptide of the invention.

The FGFR kinase inhibitor used in the afore-mentioned therapy methods can be selected from the group consisting of 5-amino-1-(2-methyl-1H-benzimidazol-5-yl)-1H-pyrazol-4-yl]-(1H-indol-2-yl)-methanone, PD173074, Pazopanib, AZD4547, Ponatinib, Dovitinib, BGJ398, E-3810, JNJ-42756493, ARQ 087, LY2874455, BAY1163877, ASP5878, E7090, ODM-203, Nintedanib, TAS-120, PRN 1109 and PRN 1371. More specifically, the FGFR kinase inhibitor can be 5-amino-1-(2-methyl-1H-benzimidazol-5-yl)-1H-pyrazol-4-yl]-(1H-indol-2-yl)-methanone.

The present invention also concerns a method of characterizing a tumor in a human subject suffering from cancer, comprising assaying protein or nucleic acids of a biopsy or fluid sample containing cancer cells or tumor circulating DNA from the subject to identify the presence or absence of a gene encoding a fusion polypeptide of the invention or an expressed fusion polypeptide of the invention. In a particular embodiment, the cancer is cholangiocarcinoma.

Finally, the invention also relates to screening methods for identifying a compound having FGFR inhibitory activity. Such a method comprises (a) culturing a cell that expresses a fusion polypeptide of the invention and whose growth is dependent on this expression in the presence or absence of a test compound and determining the level of cell proliferation; (b) comparing the proliferation level of the cultured cell in the presence and absence of the test compound; and (c) judging that the test compound has FGFR inhibitory activity when the proliferation level of the cell cultured in the presence of the test compound is lower than that of the cell cultured in the absence of the test compound. The cell utilized in the method can be a cancer cell and, more specifically, a cholangiocarcinoma cell.

BRIEF DESCRIPTION OF FIGURES

Fig. 1 shows PCR fragments amplified from cDNA from samples of two different patients, of which one has an FGFR2-VCL fusion gene (A) and the other an FGFR2-

CCDC147 fusion gene (B), and visualized after agarose gel electrophoresis. Primer pairs used for amplification listed in Table 1 and in SEQ ID NOs: 8, 9-12, 14-15): 1a: primer pair 16/17, 1b: primer pair 16/18, 2a: primer pair 16/19, 2b: primer pair 16/20, 3a: primer pair 16/21, and 3b: primer pair 16/22. Fragment size standards are shown to the left of the experimental lanes.

Fig. 2 is a graph representing the plating efficiency (%) results for Rat2 cells and derived cell pools expressing FGFR2-VCL and FGFR2-CCDC147 obtained as per Example 4.

Fig. 3 shows representative images of colonies of Rat2 cells obtained after 21 days of incubation of single cells in soft agar as per Example 4: parental cells (A), FGFR2-VCL expressing cells (B) and FGFR2-CCDC147 expressing cells (C).

Fig. 4 shows dose response curves obtained with Rat2 cells in the presence of FGFR inhibitors as per Example 5(1): parental cells (A), FGFR2-VCL-expressing cells (B) and FGFR2-CCDC147-expressing cells (C). Growth rates were compared. Figure 4 also shows a table summarizing IC50s obtained with the different inhibitors (D).

Fig. 5 shows dose response curves obtained with Rat2 cells in the presence of FGFR inhibitors as per Example 5(2): in parental cells (A), FGFR2-VCL-expressing cells (B) and FGFR2-CCDC147-expressing cells (C). Growth rates were compared. Figure 5 also shows a table summarizing IC50s obtained with the different inhibitors (D).

Fig. 6 is a graph showing primary tumor volumes measured as per Example 6 in mice inoculated with Rat2 cells (Groups 1 and 2), Rat2-FGFR2-CCDC147 cells (Groups 3 and 4) and Rat2-FGFR2-VCL cells (Groups 5 and 6), respectively. Data are displayed as means ± SEM.

Fig. 7 is a graph showing tumor volumes (A), body weight (B) and tumor weights on the last day of treatment (C) as obtained in the FGFR2-CCDC147 in vivo model of Example 7(1). Data are displayed as means ± SEM. P-values were calculated compared to the Vehicle Control using the Mann Whitney test (unpaired t-test in parentheses).

Fig. 8 is a graph showing tumor volumes (A), body weight (B) and tumor weights on the last day of treatment (C) as obtained in the FGFR2-VCL in vivo model of Example 7(2). Data are displayed as means ± SEM. P-values were calculated compared to the Vehicle

Control and between Groups 2 and 3 using the Mann Whitney test (unpaired t-test in parentheses).

Fig. 9 is a western blot analysis of HEK293T cells transiently mock-transfected or transfected with FGFR2-VCL or FGFR-CCDC147 expression construct and briefly treated with Vargatef, Compound A or vehicle as detailed in Example 8(1). Anti-myc antibody (α-myc) detected the myc-tagged fusion polypeptides, and anti-phosphotyrosine antibody (α-pY) detected phosphorylation of the fusion polypeptides.

Fig. 10 is a graph reporting results from a similar experiment as in Fig. 9, in which autophosphorylation of fusion polypeptides was quantified by a sandwich phosphotyrosine ELISA as detailed in Example 8(2). Optical density data are shown.

DETAILED DESCRIPTION OF THE INVENTION

Definitions:

In the present application, the references used for all genomic sequences or annotation of sequences are: the Genome Reference Consortium Human Build 37 (GRCh37) and Ensembl v42 assembly (Flicek, P. et al. (2014) Nucleic Acids. Res. 42; D749-55, Database issue).

"FGFR" refers to any member of the family of fibroblast growth factor receptors. The FGFR family is a member of the receptor tyrosine kinase family. Four members of the FGFR family are known, i.e., FGFR1, FGFR2, FGFR3 and FGFR4. The FGFR as referred to in the present invention may be from any origin, but preferably from a mammal and, more preferably, from a human. The most preferred FGFR is FGFR2. The chromosomal location of the human FGFR2 gene is 10q26.

"CCDC147" refers to a polypeptide known as "coiled-coil domain containing 147". The CCDC147 gene or polypeptide as referred to in the present invention may be from any origin, but preferably from a mammal and, more preferably, from a human. Expression of the polypeptide has been observed in kidney, liver, lung and blood (platelets). The chromosomal location of the human CCDC147 gene is 10q25.1. It is noted that CCDC147 now appears on Ensembl under the name CFAP58 (cilia and flagella associated protein 58).

"VCL" refers to vinculin. Vinculin is a cytoskeletal polypeptide associated with cell-cell and cell-matrix junctions. The VCL gene or polypeptide as referred to in the present invention may be from any origin, but preferably from a mammal and, more preferably, from a human. The tail of vinculin containing a five-helix bundle is known to self-associate. See Campbell, S. (2006) Abstract: P26.00007: The chromosomal location of the human VCL gene is 10q22.2.

"Wild-type" as applied to FGFR2 nucleotide sequences (i.e., nucleotide sequences encoding a wildtype FGFR2 polypeptide) refers to any known FGFR2 nucleotide sequence capable of translation into a full-length polypeptide, in particular to any of SEQ ID NOs: 16, 22, 24, 26, 28, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, or 58 (ENST00000358487, '357555, '351936, '360144, '457416, '346997, '369056, '369058, '369061, '369059, '369060, '356226, '336553, '478859, '429361). "Wild-type" as applied to FGFR2 amino acid sequences refers to any known full-length FGFR2 polypeptide sequence, in particular to any of SEQ ID NOs: 17 (or 23), 25, 27, 29, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57 or 59 (ENSP00000351276, '350166, '309878, '353262, '410294, '263451, '358052, '358054, '358057, '358055, '358056, '348559, '337665, '474011, '404219), respectively. "Wild-type" as applied to VCL nucleotide sequences (i.e., nucleotide sequences encoding a wildtype VCL polypeptide) refers to any known VCL nucleotide sequence capable of translation into a full-length polypeptide, in particular to any of SEQ ID NOs: 18, 30, 32, 34 or 60 (ENST00000211998, '372755, '417648, '436396). "Wild-type" as applied to VCL amino acid sequences refers to any known full-length VCL polypeptide sequence, in particular to any SEQ ID NOs: 19 (or 31), 33, 35 or 61 (ENSP00000211998, '361841, '411887, '415489), respectively. "Wild-type" as applied to CCDC147 nucleotide sequences (i.e., nucleotide sequences encoding a wildtype CCDC147 polypeptide) refers to any known CCDC147 nucleotide sequence capable of translation into a full-length polypeptide, in particular to SEQ ID NOs: 20, 36 or 62 (ENST00000369704, '369703). "Wild-type" as applied to CCDC147 amino acid sequences refers to any known full-length CCDC147 polypeptide sequence, in particular to SEQ ID NOs: 21 (or 37) or 63 (ENSP00000358718, '358717), respectively. In the same contexts, "mutant" refers to a sequence that differs by at least one nucleotide or one amino acid from one of the sequences of SEQ ID NOs: 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60 or 62, or SEQ ID NOs: 17,

19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61 or 63, respectively.

"Part of a polypeptide" refers to a polypeptide consisting of an arbitrary portion of the amino acid sequence of a full-length polypeptide.

"Fusion polypeptide" refers to a polypeptide in which the whole or a part of a wildtype or mutant FGFR polypeptide is fused to all or a part of a different polypeptide, at a location called a "fusion point". In the specific context of the present invention the term refers to a polypeptide in which the whole or a part of a wild-type or mutant FGFR2 polypeptide is fused to the whole or a part of a wild-type or mutant CCDC147 polypeptide or the whole or a part of a wild-type or mutant VCL polypeptide.

"Fusion gene" refers to a gene encoding a fusion polypeptide. A fusion gene also comprises a fusion point, also called genomic breakpoint.

"Cancer" generally refers to malignant neoplasm, which may be metastatic or non-metastatic. For instance, non-limiting examples of cancer that develops from epithelial tissues such as gastrointestinal tract and skin include brain tumor, skin cancer, head and neck cancer, esophageal cancer, lung cancer, stomach cancer, duodenal cancer, breast cancer, prostate cancer, cervical cancer, cancer of uterine body, pancreatic cancer, liver cancer, cholangiocarcinoma, gallbladder cancer, colorectal cancer, colon cancer, bladder cancer, and ovarian cancer. Non-limiting examples of sarcoma that develops from non-epithelial tissues (stroma) such as muscles include osteosarcoma, chondrosarcoma, rhabdomyosarcoma, leiomyosarcoma, liposarcoma, and angiosarcoma. Furthermore, non-limiting examples of hematological cancer derived from hematopoietic organs include malignant lymphoma including Hodgkin's lymphoma and non-Hodgkin's lymphoma, leukemia including acute myelocytic leukemia, chronic myelocytic leukemia, acute lymphatic leukemia, chronic lymphatic leukemia, and multiple myeloma.

By a "therapeutically effective amount" of an active agent, e.g., a drug substance, is meant an amount of the compound which, subsequent to single or multiple administration, confers a therapeutic effect on the treated subject, at a reasonable benefit/risk ratio applicable to any medical treatment. The therapeutic effect may be objective (i.e., measurable by some test or marker) or subjective (i.e., subject gives an

indication of or feels an effect). However, it is understood that effective doses will also vary depending on route of administration, as well as the possibility of co-usage with other agents. It will be understood, however, that the total daily usage of the compositions of the present invention will be decided by the attending physician within the scope of sound medical judgment. The specific therapeutically effective dose level for any particular patient will depend upon a variety of factors including the disorder being treated and the severity of the disorder; the activity of the specific active agent employed; the specific composition employed; the age, body weight, general health, sex and diet of the patient; the time of administration, route of administration, and rate of excretion of the specific active agent employed; the duration of the treatment; drugs used in combination or contemporaneously with the specific active agent employed; and like factors well known in the medical arts.

The term "pharmaceutically acceptable carrier" or "pharmaceutically acceptable vehicle" encompasses any of the standard pharmaceutical carriers, solvents, surfactants, or vehicles. Suitable pharmaceutically acceptable vehicles include aqueous vehicles and non-aqueous vehicles. Standard pharmaceutical carriers and their formulations are described in Remington's Pharmaceutical Sciences, Mack Publishing Co., Easton, PA, 19th ed. 1995.

“Specifically binding to” or “specifically hybridizing to” means that two oligo- or polynucleotides interact with one another but not detectably with any different oligo- or polynucleotide under the given conditions, or, if conditions are not given, under adequate conditions that can be identified based on knowledge in the art.

An exon is any nucleotide sequence encoded by a gene that remains present within the final mature RNA product of that gene after introns have been removed by RNA splicing. The term “exon” refers to both the DNA sequence within a gene and to the corresponding sequence in RNA transcripts and cDNAs derived therefrom. For the purposes of the present application, the numbering of exons starts with exon number 1 which contains 5’ untranslated sequences. It is noted that the cDNAs of SEQ ID NOs: 16, 18 and 20 include untranslated sequences, whereas SEQ ID NOs: 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62 as well as the subsequences of SEQ ID NOs: 64-72 only contain polypeptide-coding sequences. It is further noted that the exon annotation is performed on the basis of the longest coding

transcript found in Ensembl v42 assembly for each part of a fusion gene, i.e. SEQ ID NO: 16 for FGFR2, SEQ ID NO: 18 for VCL and SEQ ID NO: 20 for CCDC147.

As used herein, "cDNA" refers to a partial or complete copy of a gene transcript. The term is meant to encompass the latter copy, its complement as well as the double-stranded DNA consisting of both copy and complement.

The present invention relates to novel fusion polypeptides that are expressed in certain human cancer cells but not in normal cells. More specifically, it relates to fusion polypeptides comprising an FGFR2 polypeptide and a CCDC147 polypeptide or a VCL polypeptide, wherein the FGFR2 polypeptide is the whole or a part of a wildtype FGFR2 polypeptide, or the whole or a part of a mutant FGFR2 polypeptide having one or more amino acid substitutions, deletions, or insertions with respect to the wildtype polypeptide; the CCDC147 polypeptide is the whole or a part of a wildtype CCDC147 polypeptide, or the whole or a part of a mutant CCDC147 polypeptide having one or more amino acid substitutions, deletions, or insertions with respect to the wildtype polypeptide; and the VCL polypeptide is the whole or a part of a wildtype VCL polypeptide, or a part of a mutant VCL polypeptide having one or more amino acid substitutions, deletions, or insertions with respect to the wildtype polypeptide, wherein claimed fusion polypeptides are recombinant polypeptide, are isolated from cancer cells that are propagated *in vitro* or as xenografts, or are isolated from human cholangiocarcinomas. Taking into account SEQ ID NOs, the invention relates to novel fusion polypeptides comprising an FGFR2 polypeptide and a CCDC147 polypeptide or a VCL polypeptide, whereby the FGFR2 polypeptide is the whole or a part of a wildtype polypeptide having the amino acid sequence of any of SEQ ID NOs: 17, 25, 27, 29, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57 or 59, or the whole or a part of a mutant polypeptide having one or more amino acid substitutions, deletions, or insertions with respect to the wildtype polypeptide; the CCDC147 polypeptide is the whole or a part of a wildtype polypeptide having the amino acid sequence of SEQ ID NOs: 21 or 63, or the whole or a part of a mutant polypeptide having one or more amino acid substitutions, deletions, or insertions with respect to the wildtype polypeptide; and the VCL polypeptide is the whole or a part of a wildtype polypeptide having the amino acid sequence of any of SEQ ID NOs: 19, 33, 35 or 61, or the whole or a part of a mutant polypeptide having one or more amino acid substitutions, deletions, or insertions with respect to the wildtype polypeptide. Preferably, the latter

amino acid substitutions, deletions, or insertions affect (i.e. substitute, add or delete) 1-10 amino acids, more preferably 1-5 amino acids and most preferably 1-2 amino acids, in a polypeptide sequence or a fragment of a polypeptide sequence present in the fusion polypeptide. Mutant FGFR2, CCDC147 or VCL polypeptides also encompass FGFR2, CCDC147 or VCL polypeptides having 70% or more identity, preferably 80% or more identity, more preferably 90% or more identity and even more preferably 95% or more identity with the respective wild-type polypeptides or fragments thereof. Most preferably, mutant polypeptides have at least 97%, at least 98% or at least 99% identity with the respective wild-type polypeptides or fragments thereof.

The identity of an amino acid sequence (or a nucleotide sequence) relative to another can be determined using the algorithm BLAST. Karin and Altschul (1993) Proc. Natl. Acad. Sci. USA 90: 5873-7. Programs such as BLASTN and BLASTX were developed based on this algorithm. Altschul. Et al. (1990) J. Mol. Biol. 215: 403-10. To analyze nucleotide sequences according to BLASTN, parameters for score can be set at 100, and wordlength at 12. When analyzing amino acid sequences using BLASTX, score can be at 50 and wordlength at 3. Default parameters can be used when using the BLAST and Gapped BLAST programs. Specific techniques for such analyses are known in the art. Reference is made to the information on the website of the National Center for Biotechnology Information (NCBI), Basic Local Alignment Search Tool (BLAST).

The present invention also encompasses fusion polypeptides comprising an FGFR2 polypeptide and a CCDC147 polypeptide or a VCL polypeptide, wherein the FGFR2 polypeptide is the whole or a part of an FGFR2 polypeptide identifiable in a mammalian animal proteome, or the whole or a part of a mutant polypeptide that differs from the latter polypeptide by one or more amino acid substitutions, deletions, or insertions; the CCDC147 polypeptide is the whole or a part of a CCDC147 polypeptide identifiable in a mammalian animal proteome, or the whole or a part of a mutant polypeptide that differs from the latter polypeptide by one or more amino acid substitutions, deletions, or insertions; and the VCL polypeptide is the whole or a part of a VCL polypeptide identifiable in a mammalian animal proteome, or the whole or a part of a mutant polypeptide that differs from the latter polypeptide by one or more amino acid substitutions, deletions, or insertions.

The present invention further relates to polynucleotides encoding fusion polypeptides. More specifically, it relates to a polynucleotide, such as a cDNA, encoding a fusion polypeptide comprising an FGFR2 polypeptide and a CCDC147 polypeptide or a VCL polypeptide, wherein the polynucleotide combines a first polynucleotide encoding the whole or a part of a wildtype FGFR2 polypeptide or a mutant polynucleotide derived therefrom by substitution, deletion or insertion of one or more codons; and a second polynucleotide encoding the whole or a part of a wildtype CCDC147 polypeptide or VCL polypeptide or a mutant polynucleotide derived therefrom by substitution, deletion or insertion of one or more codons. Taking into account SEQ ID NOs, the present invention relates to a polynucleotide (cDNA) encoding a fusion polypeptide comprising an FGFR2 polypeptide and a CCDC147 polypeptide or a VCL polypeptide, wherein the polynucleotide combines a first polynucleotide encoding the whole or a part of a wildtype FGFR2 polypeptide, the first polynucleotide encompassing all or part of the nucleotide sequence of any of SEQ ID NOs: 16, 22, 24, 26, 28, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56 or 58, or a nucleotide sequence derived from any of SEQ ID NOs: 16, 22, 24, 26, 28, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56 or 58 by substitution, deletion or insertion of one or more codons; and a second polynucleotide encoding the whole or a part of a wildtype CCDC147 polypeptide or VCL polypeptide, the second polynucleotide encompassing all or part of the nucleotide sequence of any of SEQ ID NOs: 20, 36 or 62, or any of SEQ ID NOs: 18, 30, 32, 34 or 60, or a nucleotide sequence derived from the nucleotide sequence of any of SEQ ID NOs: 20, 36 or 62, or any of SEQ ID NOs: 18, 30, 32, 34 or 60 by substitution, deletion or insertion of one or more codons. Preferably, the latter substitutions, deletions, or insertions of codons in the nucleotide sequences encoding FGFR2, CCDC147 or VCL affect (i.e. substitute, add or delete) 1-10 amino acids, more preferably 1-5 amino acids and most preferably 1-2 amino acids in the encoded polypeptide or the fragment of the polypeptide present in the fusion polypeptide. The polynucleotides of the invention also include polynucleotides that encode FGFR2, CCDC147 or VCL polypeptides independently having 70% or more identity, preferably 80% or more identity, more preferably 90% or more identity and even more preferably 95% or more identity with the respective wild-type polypeptides or fragments thereof. Most preferably, the derived polynucleotides encode polypeptides or fragments thereof that have at least 97%, at least 98% or at least 99% identity with the respective wildtype polypeptides or fragments thereof.

Polynucleotides of the present invention can be obtained by any method. They include, e.g., all cDNAs prepared from messenger RNAs (mRNAs), DNAs derivatized from genomic DNAs, DNAs prepared by chemical synthesis, DNAs obtained by polymerase chain reaction (PCR) amplification from RNA or DNA templates as well as DNAs prepared by a combination of the latter methods. Non-genomic-type polynucleotides encoding fusion polypeptides of the present invention can be obtained by synthesis of cDNA from mRNA encoding a fusion polypeptide, by isolation of a genomic DNA fragment followed by removal of intervening sequences from the fusion polypeptide-coding region, or by chemical synthesis using a method known in the art.

To provide an example, total RNA may be prepared from cells or tissues expressing a fusion polypeptide of the invention. Total RNA may be obtained, e.g., by the guanidine-isothiocyanate method, the hot phenol method or the acid guanidinium thiocyanate-phenol-chloroform method. Messenger RNA may be selected by affinity chromatography on oligo(dT) cellulose, polyU Sepharose or the like. Using such mRNA as a template, cDNA synthesis can be carried out using a known method, e.g., using the reverse transcriptase reaction. Mol. Cell. Biol. 2: 161 (1982); Mol. Cell. Biol. 3: 280 (1983); Gene 25: 263 (1983). Subsequent to second strand synthesis, the double-stranded cDNA is inserted into a vector such as a plasmid, a phage, a cosmid or the like. The resulting library is then introduced into appropriate host cells, e.g., *E. coli*, and screened for the presence in a host cell of a vector carrying a cDNA for a fusion polypeptide of the invention using methods known in the art, e.g., colony hybridization.

The present invention also encompasses vectors carrying a polynucleotide encoding a fusion polypeptide of the invention. The vectors are not particularly limited, provided they replicate or amplify autonomously in a prokaryotic or eukaryotic host cell. Polynucleotides of the invention may be introduced into vectors using standard molecular biology techniques. Vectors include *E. coli*-derived plasmids such as pBR322, pBR325, pUC12. pUC13, pUC19, yeast (e.g., *S. cerevisiae*)-derived plasmids such as the pYC vectors or pRS shuttle vectors, and *Bacillus subtilis*-derived plasmids such as pUB110, pTP5 or pC194. Viral vectors include bacteriophage vectors such as λgt10 and λgt11, and insect or animal virus vectors derived from nuclear polyhedrosis virus, retroviruses including lentiviruses, vaccinia virus, etc.

The invention also relates to expression vectors that allow for insertion of a polynucleotide of the invention and expression of a fusion polypeptide of the invention in a prokaryotic or eukaryotic host. Suitable expression vectors include pEF-BOS (Nucleic Acids Res. 18: 5322 (1990)) and pME18S-FL (Addgene plasmid repository database, www.addgene.org, March 7, 2014). Fusion polypeptides of the present invention can also be expressed as (further) fusions with other polypeptides. Plasmid pGEX4T1 is suitable for production of a fusion polypeptide as a fusion with glutathione S-transferase sequences. Using an appropriate vector, a fusion polypeptide of the invention can also be expressed, e.g., as fusion with influenza hemagglutinin, immunoglobulin constant region, β-galactosidase, or maltose-binding protein (e.g. using a pMAL C2 vector). Fusions with various peptides are also encompassed, such as fusions with FLAG (Hopp, T.P. et al. (1988) BioTechnology 6: 1204-10), 6 x His consisting of 6 histidine residues, 10 x His, influenza hemagglutinin fragments, fragments of human c-myc, fragments of VSV-GP, fragments of p18 HIV, T7-tag, HSV-tag, E-tag, fragments of SV40 T antigen, lck tag, fragments of α-tubulin, B-tag, fragments of Protein C, Stag, StrepTag and HaloTag.

It is understood that expression vectors contain all elements required for efficient transcription and translation of a cDNA gene for a fusion polypeptide of the invention, including elements such as promoters, transcriptional and translational enhancers, start/stop codons for translation, ribosome binding site signal, transcription termination signal, polyadenylation site, signal for secretion of fusion polypeptide into the medium (or periplasmic space), etc., as well as for replication. Furthermore, vectors may contain marker genes (genes for amplification, drug resistance genes, etc.) that enable selection of transformed hosts or hosts with gene amplification. Example promoters include the *E.coli* Trp, lac, recA, IPL, Ipp and tac promoters, the yeast PH05, PGK, GAP and ADH promoters, the *B. subtilis* SL01, SP02 and penP promoters and the mammalian SV40, retroviral and heat shock promoters. Example marker genes include the dihydrofolate reductase gene, thymidine kinase gene, neomycin resistance gene, glutamate synthase gene, adenosine deaminase gene, ornithine decarboxylase gene, hygromycin-B-phosphotransferase gene, and aspartate transcarbamylase gene. A polynucleotide of the invention may be introduced into an expression vector using standard molecular biology techniques.

The present invention further relates to recombinant cells transformed with vectors including expression vectors carrying a polynucleotide of the invention. There is no particular limitation as to cell types that may be so transformed. Preferred are *E. coli* and animal, including in particular mammalian, cells. Examples of *E. coli* cells are DH5α, TB1 and HB101, of mouse cells are COP, L, C127, Sp2/0, NS-1 and NIH3T3 cells, of rat cells are PC12 and PC12h, of hamster cells are BHK and CHO, of monkey cells are COS1. COS3, COS7, CV1 and Vero, and of human cells are HeLa cells, diploid fibroblast-derived cells, myeloma cells and HepG2 cells. Methods for introducing vectors into host cells were described in Proc. Natl. Acad. Sci. USA 69: 2110 (1972); Mol. Gen. Genet. 168: 111 (1979); J. Mol. Biol. 56: 209 (1971); Proc. Natl. Acad. Sci. USA 75: 1927 (1978); J. Bacteriol. 153: 163 (1983); Virology 52: 456 (1973); Mol. Cell. Biol. 3: 2156 (1983).

Fusion polypeptide of the invention can be produced by culturing recombinant cells as described immediately before, preferably animal including mammalian animal or human cells, and, provided that the cells are capable of secreting the polypeptide, thereafter collecting the culture medium from which cells and cell debris are removed by filtration or centrifugation. Fusion polypeptide can then be purified by conventional methods such as methods based on solubility, e.g., salting out and solvent precipitation, methods based on molecular size, e.g., dialysis, ultrafiltration, gel filtration, and native and SDS-PAGE, methods based on charge, e.g., ion exchange chromatography and hydroxyapatite chromatography, affinity-based methods, e.g., affinity column chromatography, methods exploiting hydrophobicity, e.g., reverse phase high performance liquid chromatography, and methods based on isoelectric differences between polypeptides, e.g., isoelectric focusing.

When fusion polypeptide produced in the recombinant cells is accumulating in the cytoplasm/nucleoplasm or in the periplasm of walled recombinant cells, cells are collected by methods such as filtration or centrifugation. After suspension of the cells in an appropriate buffer, the cell wall or cell membrane, respectively, is disrupted by methods such as sonication, lysozyme treatment or cryolysis, and a membrane or cytoplasmic/nucleoplasmic fraction is obtained after centrifugation or filtration. Fusion polypeptide is then purified by the biochemical methods described immediately before.

The present invention also encompasses oligonucleotides that hybridize to polynucleotides of the invention. Pairs of sense and antisense oligonucleotides that are complementary to sequences within a polynucleotide of the invention are useful as primers for the amplification of the polynucleotide or parts thereof by PCR. Oligonucleotide primers that are complementary to a polynucleotide of the invention can be of any length. Preferably, they contain a sequence of at least 12 consecutive nucleotides, more preferably 12-50 consecutive nucleotides and most preferably 18-30 consecutive nucleotides. Typically, such oligonucleotides have no internal secondary structure, have a G/C content of between 40 and 60% and have a balanced distribution of G/C and A/T rich domains. When oligonucleotides are used in hybridization procedures, preference increases with the length of their sequence. Hence, oligonucleotides of > 300 consecutive nucleotides are preferred over those of > 200 consecutive nucleotides, which are preferred over oligonucleotides of > 100 consecutive nucleotides. Oligonucleotides of > 100 consecutive nucleotides are preferred over those of > 50 consecutive nucleotides which are preferred over oligonucleotides of 20 to 50 consecutive nucleotides.

Also encompassed are oligonucleotides that are complementary to a portion of an mRNA sequence encoding a fusion polypeptide of the invention. Such oligonucleotides can function as antisense oligonucleotides, ribozymes or small interfering RNAs (siRNAs). Antisense oligonucleotides bind target mRNAs or genomic DNA and inhibit their translation or transcription, respectively. Antisense oligonucleotides containing a sequence of 5-70 consecutive nucleotides are preferred over those containing 5-100 consecutive nucleotides: More preferred are oligonucleotides containing 5-50 consecutive nucleotides. Even more preferred are antisense oligonucleotides that contain 5-30 consecutive nucleotides. Antisense oligonucleotides can be modified to enhance their stability in blood, resistance to degradation or absorption in the digestive tract, or membrane permeability. Modifications of phosphate bonds include conversion of one or more bonds to phosphodiester, phosphorothioate, phosphorodithioate, methyl phosphonate, phosphoroamidate, non-phosphate bonds or phosphonothionate bonds. Ribose may be converted to 2'-fluororibose or 2'-O-methylribose. Modified nucleotide bases include 5-propynyluracil and 2-aminoadenine. siRNAs typically are double-stranded RNAs (dsRNAs) of 10-25 nucleotides in length that are capable of RNA interference. Bass (2001) Nature 411: 428-429; Elbashir et al. (2001) Nature 411: 494-

98. Like antisense oligonucleotides, siRNAs can also comprise chemically modified nucleotides and non-nucleotides. The more recently described single-stranded siRNAs are also encompassed herein. Lima, W.F. et al. (2012) Cell 150: 883-894; Yu, D. et al. (2012) Cell 150: 895-908.

In a specific embodiment, an oligonucleotide of the present invention used in hybridization procedures (i.e., as oligonucleotide probe) is complementary to a region that is composed of sequences from both fusion partners, i.e., from FGFR2 and from CCDC147 or VCL. In a more specific embodiment, it is complementary to a region comprising the fusion point. In another specific embodiment, one primer of a primer pair used for amplification of a polynucleotide of the invention or a part thereof is identical or complementary to FGFR2 sequences, whereas the other primer is identical or complementary to VCL or CCDC147 sequences.

The present invention also relates to antibodies and antigen-binding fragments that bind to any part of a fusion polypeptide of the invention. The invention is not limited to any particular type of antibody. Antibodies of the invention may be any antibody, from any mammalian source, and either polyclonal or monoclonal. Methods for the production of mammalian monoclonal antibodies, e.g., mouse monoclonal antibodies, have long been established and are widely practiced by persons skilled in the art. Kohler and Milstein (1981) Meth. Enzymol. 73: 3-46. Preferred antibodies for administration to human subjects are chimeric antibodies, humanized antibodies and human antibodies. They are preferred because they do not provoke, or provoke less, human host immune responses as opposed to, e.g., antibodies from a mammalian animal such as a mouse. Methods for preparing chimeric antibodies containing human constant regions and variable regions from a mammalian animal are well known in the art. Carl, A.K. et al. "Therapeutic monoclonal antibody", published in the U.K. by McMillan Publishers LTD., 1990. The same is true for humanized antibodies. Patent publications EP 125023; WO 96/02576; WO 98/13388; EP 239400; WO 96/02576. Techniques for directly producing humanized antibodies by introducing complementarity-determining regions (CDR) sequences from an animal antibody gene into a human antibody template gene have also been described. There exist several well-known approaches for obtaining human antibodies. Human monoclonal antibodies can be produced by immunization of human lymphocytes *in vitro* followed by fusion to a human lymphoblastoid cell line. Antibodies can then be produced from the resulting fusion cells by biotechnological methods. Borrebaek et al.

(1988) Proc Natl. Acad. Sci. USA 85: 3995-9. A human antibody can also be obtained from immunization of a transgenic animal carrying an entire repertoire of human antibody genes. International publications nos. WO 2003/12227; WO 92/03918; WO 94/02602; WO 94/25585; WO 96/34096 and WO 96/33735. In another approach, human B cells expressing an antibody against a fusion polypeptide of the invention can be selected using a suitable approach such as flow cytometry. The nucleotide sequence of the antibody can then be determined. Jin et al. (2009) Nat. Med. 15: 1088-92; Scheid et al. (2009) Nature 458: 636-40; Wrammert et al. (2008) Nature 453: 667-72; Tiller et al. (2008) J. Immunol. Meth. 329: 112-24. This information is subsequently used to obtain DNA sequences encoding the antibody, construct an appropriate expression vector and produce the antibody by biotechnological methods. International publications nos. WO 92/01047; WO 92/20791; WO 93/06213; WO 93/11236; WO 93/19172; WO 95/01438 and WO 95/15388. Another approach involves panning of a human antibody phage display library, in which single chain human antibodies (human scFv) are displayed on the surfaces of bacteriophages. Subsequent to determination of the nucleotide sequence encoding a selected scFv, a complete antibody gene can be constructed and expressed in a suitable producer cell. International publications nos. WO 92/01047; WO 92/20791; WO 93/06213; WO 93/11236; WO 93/19172; WO 95/01438 and WO 95/15388.

The antibodies of the invention include divalent antibodies as represented by IgG as well as monovalent antibodies as represented by IgM. Bi-specific antibodies binding to two different antigens are also encompassed. Also encompassed are chimeras of an antibody and a toxic product or polypeptide. Antibodies of the invention also include antigen-binding fragments such as minibodies. Minibodies comprise only a portion of an antibody, typically including six CDR sequences. Specific examples of minibodies include Fab, Fab', F(ab')2, Fv, scFv (single chain Fv), diabodies, and sc(Fv)2 (single chain (Fv)2), and multimers thereof. For the production of antigen-binding molecule fragments, see Co *et al.*, (1994) J. Immunol. 152: 2968-76; Better and Horwitz (1989) Meth. Enzymol. 178: 476-96; Plueckthun and Skerra (1989) Meth. Enzymol. 178: 476-96; Lamoyi (1989) Meth. Enzymol. 121: 652-63; Rousseaux et al. (1989) Meth. Enzymol. 121: 663-69; Bird et al. (1991) TIBTECH 9: 132-7. For diabodies see Holliger et al. (1993) Proc. Natl. Acad. Sci. USA 90: 6444-8; patent publications EP 404,097 and WO 93/11161. For scFv antibodies, see Huston et al., (1988) Proc. Natl. Acad. Sci. U.S.A. 85, 5879-83; Plickthun "The Pharmacology of Monoclonal Antibodies" Vol. 113, eds.,

Resenburg and Moore, Springer Verlag, New York, pp. 35 269-315 (1994). sc(Fv)2 is a single-chain minibody produced by linking two VHs and two VLs using linkers and such (Hudson et al. (1999) J Immunol. Methods 231: 177-89). Also encompassed are antibody-drug conjugates (ADC), wherein the antibodies of the present invention are linked to cytotoxic agents, e.g. via linkers.

In specific embodiments, the antibodies and antigen-binding fragments of the present invention bind to an epitope that is composed of sequences from both fusion partners, i.e., from FGFR2 and from CCDC147 or VCL. In further specific embodiments, they bind to an epitope that comprises the fusion point.

The present invention further relates to the treatment of diseases or conditions, notably including cancer, in which affected tissues express a fusion polypeptide of the invention. Treatment can involve a regimen of administration of a therapeutically effective amount of an active agent. An active agent can be an antibody or antigen-binding fragment that binds to a fusion polypeptide of the invention, an antisense RNA or siRNA molecule directed towards fusion polypeptide mRNA (or gene) or an inhibitor of the kinase activity of the fusion polypeptide.

The present invention also concerns pharmaceutical compositions comprising an antibody or antigen-binding fragment that binds to a fusion polypeptide of the invention. For example, anti-FGFR2 antibodies include BAY1179470 (Kopitz C. et al. (2014) Cancer Res 74 (Suppl. 19) 7445 - Abstract DDT02-01) and FPA144 (Gemo AT. et al. (2014) Cancer Res 74 (Suppl. 19) - Abstract 5446). ADC directed against FGFR2 includes BAY1187982 (Sommer A. et al. (2014) Cancer Res. 74 (Suppl. 19) - Abstract 4491). The latter compositions will be preferentially administered parenterally, but transnasal, transpulmonary or transdermal delivery is also envisaged. The pharmaceutical compositions may contain any conventional non-toxic pharmaceutically-acceptable carriers, adjuvants or vehicles. In some cases, the pH of the formulation may be adjusted with pharmaceutically acceptable acids, bases or buffers to enhance the stability of the formulated agent or its delivery form. The term parenteral as used herein includes subcutaneous, intracutaneous, intravenous, intramuscular, intraarticular, intraarterial, intrasynovial, intrasternal, intrathecal, intralesional and intracranial injection or infusion techniques.

Injectable preparations, for example, sterile injectable aqueous or oleaginous suspensions, may be formulated according to the known art using suitable dispersing or wetting agents and suspending agents. The sterile injectable preparation may also be a sterile injectable solution, suspension or emulsion in a nontoxic parenterally acceptable diluent or solvent, for example, as a solution in 1,3-butanediol. Among the acceptable vehicles and solvents that may be employed are water, Ringer's solution, U.S.P. and isotonic sodium chloride solution. In addition, sterile, fixed oils are conventionally employed as a solvent or suspending medium. For this purpose any bland fixed oil can be employed including synthetic mono- or diglycerides. In addition, fatty acids such as oleic acid are used in the preparation of injectables. The injectable formulations can be sterilized, for example, by filtration through a bacterial-retaining filter, or by incorporating sterilizing agents in the form of sterile solid compositions which can be dissolved or dispersed in sterile water or other sterile injectable medium prior to use.

The dose of a pharmaceutical composition containing an antigen-binding molecule may be, for example, from 0.01 to 30 mg/kg for each administration. Alternatively, the cumulative dose may be, for example, from 0.001 to 100,000 mg per subject. However, the present invention is not limited by the numeric values described above. The doses and administration methods vary depending on the subject's weight, age, symptoms, and such. Those skilled in the art can set appropriate doses and administration methods in consideration of the factors described above.

The present invention also relates to pharmaceutical compositions comprising an antisense RNA or siRNA molecule directed towards an mRNA encoding a fusion polypeptide of the invention. In specific embodiments, the antisense RNA or siRNA molecule is directed towards a sequence that is composed of sequences from both fusion partners, i.e., from FGFR2 and from CCDC147 or VCL, or even more specifically to a sequence encompassing the fusion point. Pharmaceutical compositions containing dsRNAs were described, e.g., in EP patent nos. 1144623 and 1214945, and US patent no. 8'546'143. The same principles are taken to also relate to antisense RNA or to the more recently discovered single stranded siRNAs (ss siRNAs). Juliano, R. et al. (2008) Nucleic Acids Res. 36: 4158-71; Lima et al. (2012); Yu et al. (2012).

Pharmaceutical compositions containing dsRNAs may be administered by any means known in the art including, but not limited to oral or parenteral routes, including

intravenous, intramuscular, intraperitoneal, subcutaneous, transdermal, airway (aerosol), rectal, vaginal and topical (including buccal and sublingual) administration. In preferred embodiments, the pharmaceutical compositions are administered by intravenous or intraparenteral infusion or injection.

For oral administration, the pharmaceutical compositions containing dsRNAs will generally be in the form of tablets or capsules, as a powder or granules, or as an aqueous solution or suspension. Tablets for oral use may include the active ingredients mixed with pharmaceutically acceptable excipients such as inert diluents, disintegrating agents, binding agents, lubricating agents, sweetening agents, flavoring agents, coloring agents and preservatives. Suitable inert diluents include sodium and calcium carbonate, sodium and calcium phosphate, and lactose, while corn starch and alginic acid are suitable disintegrating agents. Binding agents may include starch and gelatin, while the lubricating agent, if present, will generally be magnesium stearate, stearic acid or talc. If desired, the tablets may be coated with a material such as glyceryl monostearate or glyceryl distearate, to delay absorption in the gastrointestinal tract. Capsules for oral use include hard gelatin capsules in which the active ingredient is mixed with a solid diluent, and soft gelatin capsules wherein the active ingredients is mixed with water or an oil such as peanut oil, liquid paraffin or olive oil.

For intramuscular, intraperitoneal, subcutaneous and intravenous use, the pharmaceutical compositions containing dsRNAs will generally be in the form of sterile aqueous solutions or suspensions, buffered to an appropriate pH and isotonicity. Suitable aqueous vehicles include Ringer's solution and isotonic sodium chloride. In a preferred embodiment, the carrier consists exclusively of an aqueous buffer. In this context, "exclusively" means no auxiliary agents or encapsulating substances are present which might affect or mediate uptake of dsRNA in the cells that express the target gene. Although microinjection, lipofection, viruses, viroids, capsids, capsoids, or other auxiliary agents may be required to efficiently introduce dsRNA into cell cultures, these methods and agents are not required for uptake of dsRNA *in vivo*. It is observed that the same is true for antisense RNAs and ss siRNAs. Juliano et al. (2008); Davidson, B.L. & Monteys, A.M. (2012) Cell 150: 873-5.

Pharmaceutical compositions containing dsRNAs in the form of aqueous suspensions may include suspending agents such as cellulose derivatives, sodium alginate, polyvinyl-pyrrolidone and gum tragacanth, and a wetting agent such as lecithin. Suitable preservatives for aqueous suspensions include ethyl and n-propyl p-hydroxybenzoate.

The pharmaceutical compositions containing dsRNAs also include encapsulated formulations to protect the dsRNA against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. Suitable polymer materials can be obtained, e.g., from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Pat. No. 4,522,811; international patent publication WO 91/06309; and European patent publication EP-A-43075, all of which are incorporated by reference herein.

The present invention further relates to pharmaceutical compositions comprising an inhibitor of the FGFR tyrosine kinase activity of a fusion polypeptide of the invention. Any inhibitor of FGFR kinase activity may be employed. Inhibitors include the inhibitory aminopyrazole derivatives and their pharmaceutically acceptable salts described in international patent publication WO 2011/016528, in particular 5-amino-1-(2-methyl-1H-benzimidazol-5-yl)-1H-pyrazol-4-yl]-(1H-indol-2-yl)-methanone (CAS 1265229-25-1, herein referred to as Compound A). They further include inhibitors PD173074 (Mohammadi et al. (1998) EMBO J. 17: 5896-904), Pazopanib (Harris et al. (2009) J. Med. Chem. 51: 4632-40; Keisner and Shah (2011) Drugs 71: 443-54), AZD4547 (Gavine et al. (2012) Cancer Res. 72: 2045-56), Ponatinib (or AP24534) (Huang et al. (2010) J. Med. Chem. 53: 4701-19), Dovitinib (Trudel et al. (2005) Blood 105: 2941-8; Man et al. (2014) J. Cell. Mol. Med. 18: 143-55), BGJ398 (Guagnano et al. (2011) J. Med. Chem. 54: 7066-83), E-3810 also known as Lucitanib (Bello et al. (2011) Cancer Res. 71: 1396-405), JNJ-42756493 (Squires et al. (2008) AACR Abstract 1545), ARQ 087 (Yu et al. (2011) Cancer Res. 71 (Suppl. 1) 3671), LY2874455 (Zhao G et al. Mol Cancer Ther. (2011) Nov;10(11):2200-10), BAY1163877 (Heroult et al. (2014) Cancer

Res. 74 (Suppl. 19) - Abstract 1739), ASP5878 (73rd Annual Meeting of the Japanese Cancer Association (2014) - Abstract/Poster 1411), E7090 (Saori Watanabe Miyano et al. (2015) AACR Abstract 770), ODM-203 (Holmström T. et al. 26th ENA Symposium (2014) Eur. J. Cancer 50(S6):142 - Abstract 432), Nintedanib (Roth GJ et al. J Med Chem. (2015) Feb 12;58(3):1053-63), TAS-120 (Ochiiwa, H. et al. (2013) AACR; Mol. Cancer Ther. 12(11 Suppl) Abstract A270), PRN 1109 and PRN 1371 (both in: Phan VT. et al. 26th ENA Symposium (2014) Eur. J. Cancer 50(S6):157 - Abstract 483).

A particularly preferred inhibitor is Compound A.

The pharmaceutical compositions of this invention comprising an inhibitor of the FGFR tyrosine kinase activity (also referred to below as "drug substance") may be administered orally, parenterally, by inhalation spray, topically, rectally, nasally, buccally, vaginally or via an implanted reservoir, preferably by oral administration or administration by injection. The pharmaceutical compositions may contain any conventional non-toxic pharmaceutically acceptable carriers, adjuvants or vehicles. In some cases, the pH of the formulation may be adjusted with pharmaceutically acceptable acids, bases or buffers to enhance the stability of the formulated compound or its delivery form. The term parenteral as used herein includes subcutaneous, intracutaneous, intravenous, intramuscular, intraarticular, intraarterial, intrasynovial, intrasternal, intrathecal, intralesional and intracranial injection or infusion techniques.

Liquid dosage forms for oral administration include pharmaceutically acceptable emulsions, microemulsions, solutions, suspensions, syrups and elixirs. In addition to drug substance, the liquid dosage forms may contain inert diluents commonly used in the art such as, for example, water or other solvents, solubilizing agents and emulsifiers such as ethyl alcohol, isopropyl alcohol, ethyl carbonate, ethyl acetate, benzyl alcohol, benzyl benzoate, propylene glycol, 1,3-butylene glycol, dimethylformamide, oils (in particular, cottonseed, groundnut, corn, germ, olive, castor, and sesame oils), glycerol, tetrahydrofurfuryl alcohol, polyethylene glycols and fatty acid esters of sorbitan, and mixtures thereof. Besides inert diluents, the oral compositions can also include adjuvants such as wetting agents, emulsifying and suspending agents, sweetening, flavoring, and perfuming agents.

Injectable preparations, for example, sterile injectable aqueous or oleaginous

suspensions, may be formulated according to the known art using suitable dispersing or wetting agents and suspending agents. The sterile injectable preparation may also be a sterile injectable solution, suspension or emulsion in a nontoxic parenterally acceptable diluent or solvent, for example, as a solution in 1,3-butanediol. Among the acceptable vehicles and solvents that may be employed are water, Ringer's solution, U.S.P. and isotonic sodium chloride solution. In addition, sterile, fixed oils are conventionally employed as a solvent or suspending medium. For this purpose any bland fixed oil can be employed including synthetic mono- or diglycerides. In addition, fatty acids such as oleic acid are used in the preparation of injectables.

The injectable formulations can be sterilized, for example, by filtration through a bacterial-retaining filter, or by incorporating a drug substance in the form of a sterile solid composition which can be dissolved or dispersed in sterile water or other sterile injectable medium prior to use.

In order to prolong the effect of a drug substance, it is often desirable to slow the absorption of the drug substance from subcutaneous or intramuscular injection. This may be accomplished by the use of a liquid suspension of crystalline or amorphous material with poor water solubility. The rate of absorption of the drug substance then depends upon its rate of dissolution, which, in turn, may depend upon crystal size and crystalline form. Alternatively, delayed absorption of a parenterally administered drug form is accomplished by dissolving or suspending the drug substance in an oil vehicle. Injectable depot forms are made by forming microencapsule matrices of the drug substance in biodegradable polymers such as polylactide-polyglycolide. Depending upon the ratio of drug substance to polymer and the nature of the particular polymer employed, the rate of drug release can be controlled. Examples of other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot injectable formulations are also prepared by entrapping the drug substance in liposomes or microemulsions that are compatible with body tissues.

Compositions for rectal or vaginal administration are preferably suppositories which can be prepared by mixing drug substance with suitable non-irritating excipients or carriers such as cocoa butter, polyethylene glycol or a suppository wax which are solid at ambient temperature but liquid at body temperature and therefore melt in the rectum or

vaginal cavity and release the active compound.

Solid dosage forms for oral administration include capsules, tablets, pills, powders and granules. In such solid dosage forms, drug substance is mixed with at least one inert, pharmaceutically acceptable excipient or carrier such as sodium citrate or dicalcium phosphate and/or: a) fillers or extenders such as starches, lactose, sucrose, glucose, mannitol, and silicic acid, b) binders such as, for example, carboxymethylcellulose, alginates, gelatin, polyvinylpyrrolidone, sucrose, and acacia, c) humectants such as glycerol, d) disintegrating agents such as agar-agar, calcium carbonate, potato or tapioca starch, alginic acid, certain silicates, and sodium carbonate, e) solution retarding agents such as paraffin, f) absorption accelerators such as quaternary ammonium compounds, g) wetting agents such as, for example, cetyl alcohol and glycerol monostearate, h) absorbents such as kaolin and bentonite clay, and/or i) lubricants such as talc, calcium stearate, magnesium stearate, solid polyethylene glycols, sodium lauryl sulfate, and mixtures thereof. In the case of capsules, tablets and pills, the dosage form may also comprise buffering agents.

Solid compositions of a similar type may also be employed as fillers in soft and hard-filled gelatin capsules using such excipients as lactose or milk sugar as well as high molecular weight polyethylene glycols and the like.

The solid dosage forms of tablets, dragees, capsules, pills, and granules can be prepared with coatings and shells such as enteric coatings and other coatings well known in the pharmaceutical formulating art. They may optionally contain opacifying agents and can also be of a composition that they release the active ingredient(s) only, or preferentially, in a certain part of the intestinal tract, optionally, in a delayed manner. Examples of embedding compositions that can be used include polymeric substances and waxes.

Dosage forms for topical or transdermal administration of drug substance include ointments, pastes, creams, lotions, gels, powders, solutions, sprays, inhalants or patches. Drug substance is admixed under sterile conditions with a pharmaceutically acceptable carrier and any needed preservatives or buffers as may be required. Ophthalmic formulations, ear drops, eye ointments, powders and solutions are also

contemplated as being within the scope of this invention.

The ointments, pastes, creams and gels may contain, in addition to drug substance, excipients such as animal and vegetable fats, oils, waxes, paraffins, starch, tragacanth, cellulose derivatives, polyethylene glycols, silicones, bentonites, silicic acid, talc and zinc oxide, or mixtures thereof.

Powders and sprays can contain, in addition to the drug substance, excipients such as lactose, talc, silicic acid, aluminum hydroxide, calcium silicates and polyamide powder, or mixtures of these substances. Sprays can additionally contain customary propellants such as chlorofluorohydrocarbons or environmentally friendlier propellents such as hydrofluoroalkane, C3-C6 light saturated hydrocarbons, dimethyl ether, and the like.

Transdermal patches have the added advantage of providing controlled delivery of drug substance to the body. Such dosage forms can be made by dissolving or dispensing the drug substance in the proper medium. Absorption enhancers can also be used to increase the flux of the drug substance across the skin. The rate can be controlled by either providing a rate-controlling membrane or by dispersing the compound in a polymer matrix or gel.

For pulmonary delivery, a pharmaceutical composition of the invention is formulated and administered to the patient in solid or liquid particulate form by direct administration e.g., inhalation into the respiratory system. Solid or liquid particulate forms of the drug substance prepared for practicing the present invention include particles of respirable size: that is, particles of a size sufficiently small to pass through the mouth and larynx upon inhalation and into the bronchi and alveoli of the lungs. Delivery of aerosolized therapeutics, particularly aerosolized antibiotics, is known in the art (see, for example US Pat. No. 5,767,068 to VanDevanter et al., US Pat. No. 5,508,269 to Smith et al, and international patent publication WO 98/43650, all of which are incorporated herein by reference). A discussion of pulmonary delivery of antibiotics is also found in US Pat. No. 6,014,969, incorporated herein by reference.

In general, treatment regimens according to the present invention comprise administration to a human subject in need of such treatment from 0.1 mg to 1000 mg of

drug substance (i.e., an inhibitor of FGFR tyrosine kinase activity) per day in single or multiple doses. Single dose compositions may contain such amounts or submultiples thereof to make up the daily dose.

The drug substance can, for example, be administered by injection, intravenously, intraarterially, subdermally, intraperitoneally, intramuscularly, or subcutaneously; or orally, buccally, nasally, transmucosally, topically, in an ophthalmic preparation, or by inhalation, every 4 to 120 hours, or according to the requirements of the particular drug substance. The methods herein contemplate administration of an effective amount of drug substance or pharmaceutical composition comprising the drug substance to achieve the desired or stated effect. Typically, the pharmaceutical compositions will be administered from 1 to 6 times per day or alternatively, as a continuous infusion. Such administration can be used as a chronic or acute therapy. The amount of drug substance that may be combined with pharmaceutically acceptable excipients or carriers to produce a single dosage form will vary depending on the particular mode of administration and, possibly, on the subject treated. A typical preparation will contain from 5% to 95% drug substance (w/w). Alternatively, such preparations may contain from 20% to 80% drug substance. Lower or higher doses than those recited above may be required. Specific dosage and treatment regimens for any particular subject will depend upon a variety of factors, including the activity of the specific drug substance employed, the age, body weight, general health status, sex, diet, time of administration, rate of excretion, drug combination, the severity and course of the disease, condition or symptoms, the subject's disposition to the disease, condition or symptoms, and the judgment of the treating physician.

The present invention also relates to methods for detecting a fusion polypeptide of the invention or a polynucleotide encoding the fusion polypeptide in a sample from a human or animal subject, e.g., tumor tissue, normal tissue, and various body fluid specimens containing cancer or normal cells or tumor circulating DNA (blood, serum, urine, saliva, etc.).

Fusion polypeptides of the invention typically can be detected by contacting a sample from the subject with one of the above-described antibodies or antigen-binding fragments and then detecting the presence or absence of a reaction product. The step of detecting the reaction product may be carried out with any suitable immunoassay.

Immunoassays carried out in accordance with the present invention may be homogeneous assays or heterogeneous assays. In a homogeneous assay, the immunological reaction usually involves the specific fusion polypeptide antibody or antigen-binding fragment, a labeled analyte, and the sample of interest. The signal arising from the label is modified, directly or indirectly, upon the binding of the antibody/ antigen-binding fragment to the labeled analyte. Both the immunological reaction and detection of the extent thereof can be carried out in a homogeneous solution. Immunochemical labels which may be employed include free radicals, radioisotopes, fluorescent dyes, enzymes, bacteriophages, or coenzymes.

In a heterogeneous assay approach, the reagents are usually the sample, the antibody/ antigen-binding fragment, and means for producing a detectable signal. Samples as described above may be used. The antibody/ antigen-binding fragment can be immobilized on a support, such as a bead (such as protein A agarose, protein G agarose, latex, polystyrene, magnetic or paramagnetic beads), plate or slide, and contacted with the specimen suspected of containing the antigen in a liquid phase. The support is then separated from the liquid phase and either the support phase or the liquid phase is examined for a detectable signal employing means for producing such signal. The signal is related to the presence of the analyte in the sample. Means for producing a detectable signal include the use of radioactive labels, fluorescent labels, or enzyme labels. For example, if the antigen to be detected contains a second binding site, an antibody which binds to that site can be conjugated to a detectable group and added to the liquid phase reaction solution before the separation step. The presence of the detectable group on the solid support indicates the presence of the antigen in the test sample. Examples of suitable immunoassays are immunoblotting, immunoprecipitation, immunofluorescence methods, chemiluminescence methods, electrochemiluminescence or enzyme-linked immunoassays.

Those skilled in the art will be familiar with numerous specific immunoassay formats and variations thereof which may be useful for carrying out the method disclosed herein. See generally E. Maggio, Enzyme-Immunoassay, (1980) (CRC Press, Inc., Boca Raton, Fla.); see also US Pat. No. 4,727,022 to Skold et al. titled "Methods for Modulating Ligand-Receptor Interactions and their Application", US Pat. No. 4,659,678 to Forrest et al. titled "Immunoassay of Antigens", US Pat. No. 4,376,110 to David et al., titled "Immunometric Assays Using Monoclonal Antibodies", US Pat. No. 4,275,149 to Litman

et al., titled "Macromolecular Environment Control in Specific Receptor Assays," US Pat. No. 4,233,402 to Maggio et al., titled "Reagents and Method Employing Channeling" and US Pat. No. 4,230,797 to Boguslaski et al., titled "Heterogenous Specific Binding Assay Employing a Coenzyme as Label".

Antibodies can be conjugated to a solid support suitable for a diagnostic assay (e.g., beads such as protein A or protein G agarose, microspheres, plates, slides or wells formed from materials such as latex or polystyrene) in accordance with known techniques, such as passive binding. Antibodies as described herein may likewise be conjugated to detectable labels or groups such as radiolabels (e.g., ^{35}S, ^{125}I, ^{131}I), enzyme labels (e.g., horseradish peroxidase, alkaline phosphatase), and fluorescent labels (e.g., fluorescein, Alexa, green fluorescent protein) in accordance with known techniques.

The presence or absence of a gene or mRNA encoding a fusion polypeptide of the present invention in a sample from a subject can be tested and determined, for example, by conventional methods using the above-described various oligonucleotides (a pair of oligonucleotide primers, oligonucleotide probes, etc.) of the present invention and mRNA, cDNA prepared using mRNA as a template, genomic DNA, or such in a sample (tumor tissue, normal tissue, and various body fluid specimens containing cancer or normal cells or circulating nucleic acids (blood, serum, urine, saliva, etc.)) collected from a subject. Such gene analysis methods include, for example, Northern blotting as well as numerous techniques enumerated below:

(1) Polynucleotide-based detection methods (i.e., see US Pat. nos. 5,310,625, 5,322,770, 5,561,058, 5,641,864, and 5,693,517; see also Myers and Sigua, Amplification of RNA: High-temperature reverse transcription and DNA amplification with Thermus thermophilus DNA polymerase. In: M.A. Innis, D.H. Gelfand and J.J. Sninsky, Eds., PCR Strategies, Academic Press, San Diego (1995), pp. 58-68, DNA sequencing methods (i.e., Sequencing methods by PE Biosystems (Foster City, CA); see Sanger et al. (1977) Proc. Natl. Acad. Sci. USA 74: 5463-5467).

(2) Amplification-based identification methods (i.e., US Pat. nos. 4,683,195, 4,683,202 and 4,965,188; see PCR Applications (1999) Innis et al., eds., Academic Press, San Diego; PCR Strategies (1995) Innis et al., eds., Academic Press, San Diego; PCR

Protocols (1990) Innis et al., eds., Academic Press, San Diego; and PCR Technology (1989) Erlich, ed., Stockton Press, New York, NY).

(3) Ligase chain reaction (Wu and Wallace (1988) Genomics 4: 560-569); the strand displacement assay (Walker et al. (1992) Proc. Natl. Acad. Sci. USA 89: 392-396 and Nucleic Acids Res. 20: 1691-1696; and US Pat. No. 5,455,166); and several transcription-based amplification systems, including the methods described in US Pat. nos. 5,437,990; 5,409,818; and 5,399,491; the transcription amplification system (TAS) (Kwoh et al. (1989) Proc. Natl. Acad. Sci USA 86: 1173-1177); and self-sustained sequence replication (3SR) (Guetelli et al. (1990) Proc. Natl. Acad. Sci USA 87: 1874-1878 and WO 1992/08800).

(4) Sequence-specific amplification or primer extension methods (i.e., US Pat. nos. 5,137,806; 5,595,890; 5,639,611 and 4,851,331).

(5) Kinetic PCR methods (i.e., Higuchi et al. (1992) Bio/Technology 10: 413-417; Higuchi et al. (1993) Bio/Technology 11: 1026-1030; Higuchi and Watson, in PCR Applications, supra, chapter 16; US Pat. nos. 5,994,056; EP patent publications 487,218 and 512,334).

(6) Probe-based methods that rely on the difference in stability of hybridization duplexes formed between the probe and the nucleic acid sequences in the fusion area between wild-type genes or transcripts and fusion genes or transcripts that differ in the degree of complementarity (i.e., Conner et al. (1983) Proc. Natl. Acad. Sci. USA 80: 278-282 and US Pat. nos. 5,468,613; 5,604,099; 5,310,893; 5,451,512; 5,468,613 and 5,604,099).

(7) Methods based on massive parallel sequencing of cDNA libraries. An example method is disclosed under Example 3.

Preferably, levels of expression of the fusion genes of the present invention are detected by real-time PCR, as described further in international patent publication WO 03/048377.

The present invention also encompasses kits for detecting the presence of fusion polypeptides of the invention or of genes and transcripts encoding such fusion polypeptides. Detection kits of the present invention may contain above-described antibodies or antigen-binding fragments that bind to a fusion polypeptide of the present invention. The kits may also contain, depending on the purpose of each immunoassay

described above, various detection reagents (enzymes, substrates, etc.) and instruction manuals. Other detection kits of the present invention may contain above-described various oligonucleotides of the present invention (a pair of oligonucleotide primers, oligonucleotide probes, etc.) that hybridize to mRNA encoding a fusion polypeptide of the present invention, cDNA prepared using the mRNA as a template, or genomic DNA. The kits may further contain, depending on the exact method of gene analysis employed), various reagents (enzymes, other oligonucleotides, nucleic acid, reaction buffer, etc.) and instruction manuals.

The discovery of the fusion polypeptides of the invention provides the basis for new diagnostic and therapeutic methods. The fusion polypeptides of the invention are likely to function as drivers of cancer growth. Therefore, the presence of a fusion polypeptide of the invention or the presence of a gene encoding a fusion polypeptide of the invention or its transcript in biopsied or fluid material obtained from a subject is taken to be indicative of an increased susceptibility of the subject to develop a cancerous growth or to the undetected presence in the subject of a cancerous growth. Therefore, the invention also encompasses a method for determining the susceptibility of a subject to cancer or the presence of a previously undetected cancer comprising the steps of (a) obtaining a tissue sample or a fluid sample (blood, serum, urine, saliva, etc.) from a subject to be tested, (b) determining the presence of a fusion polypeptide of the invention or of a polynucleotide sequence encoding such polypeptide in the subject's tissue or fluid sample using the methods and kits described supra, and (c) making a determination of increased susceptibility to cancer or of the likely presence of a cancer based on the positive identification of the fusion polypeptide or of the nucleotide sequence coding for the fusion polypeptide.

The presence of a fusion polypeptide of the invention or of a gene encoding a fusion polypeptide of the invention or its transcript in a tumor sample from a subject is taken to indicate that the growth of the patient's tumor would be inhibited by a therapy that results in an effective inhibition of the FGFR kinase activity of the fusion polypeptide or elimination of the fusion polypeptide. Therefore, the invention also relates to a method of personalized cancer therapy, comprising the steps of (a) taking a biopsy containing cancer cells from a subject suffering from cancer or exhibiting a potentially cancerous tumor, (b) determining, using the methods and kits being described herein, whether the cells in the biopsy contain a gene encoding a fusion polypeptide of the invention or a

transcript of such a gene, or express the fusion polypeptide, (c) selecting the subject whose biopsy could be determined to contain the gene for or express the fusion polypeptide for the treatment of step d; and (d) subjecting the selected subject to a therapeutic regimen that comprises administration of a pharmaceutical composition comprising an FGFR kinase inhibitor. Alternatively, the pharmaceutical composition can contain an antibody or antigen-binding fragment that binds to the fusion polypeptide and disrupts its biochemical function or results in its elimination (or immune system-mediated elimination of entire target cells). In yet other related methods (that share steps a-c), step (d) involves administration of a pharmaceutical composition comprising an antisense oligonucleotide that targets mRNA encoding the fusion polypeptide, inhibiting or impairing its translation, or a pharmaceutical composition comprising an siRNA directed towards mRNA encoding the fusion polypeptide, causing cleavage of the mRNA and its subsequent elimination. In specific embodiments of such methods, the antibody or antigen-binding fragment is directed to a sequence of the fusion polypeptide that comprises sequences from both fusion partners, i.e., from FGFR2 and from CCDC147 or VCL, or even more specifically to a sequence that comprises the fusion point. In other specific embodiments of such methods, the RNA or siRNA is directed to a sequence of the fusion gene comprising sequences from both fusion partners, i.e., from FGFR2 and from CCDC147 or VCL, or even more specifically to a sequence that comprises the fusion point.

Recitation of ranges of values herein are merely intended to serve as a shorthand method of referring individually to each separate value falling within the range, unless otherwise indicated herein, and each separate value is incorporated into the specification as if it were individually recited herein. Unless otherwise stated, all exact values provided herein are representative of corresponding approximate values (e. g., all exact exemplary values provided with respect to a particular factor or measurement can be considered to also provide a corresponding approximate measurement, modified by "about," where appropriate).

The use of any and all examples, or exemplary language (e.g., "such as") provided herein is intended merely to better illuminate the invention and does not pose a limitation on the scope of the invention unless otherwise indicated.

The citation and incorporation of patent documents herein is done for convenience only

and does not reflect any view of the validity, patentability and/or enforceability of such patent documents. The description herein of any aspect or embodiment of the invention using terms such as reference to an element or elements is intended to provide support for a similar aspect or embodiment of the invention that "consists of'," "consists essentially of" or "substantially comprises" that particular element or elements, unless otherwise stated or clearly contradicted by context (e. g. , a composition described herein as comprising a particular element should be understood as also describing a composition consisting of that element, unless otherwise stated or clearly contradicted by context).

This invention includes all modifications and equivalents of the subject matter recited in the aspects or claims presented herein to the maximum extent permitted by applicable law.

All publications and patent documents cited in this specification are herein incorporated by reference in their entireties as if each individual publication or patent document were specifically and individually indicated to be incorporated by reference.

Although the present invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to one of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

EXAMPLES

Example 1: Identification of fusion genes of the invention

(1) Total RNA extraction

Total RNA was extracted from two macro-dissected 10 µm thick sections from formalin-fixed, paraffin-embedded tissue obtained from biopsies of human cholangiocarcinomas, using the High Pure FFPET RNA Isolation Kit of Roche (product no. 06650775001) according to the manufacturer's instructions. The procedure involved lysis of the deparaffinized tissue using the Roche proprietary RNA tissue lysis buffer supplemented with SDS and incubation with Proteinase K. In the presence of chaotropic salts, the RNA was specifically bound to the glass fibers of the High Pure Filter Tube. Bound RNA was incubated with DNase and purified in a series of rapid wash-and-spin steps and then eluted in water. RNA concentration was determined by absorbance using a NanoDrop spectrophotometer (Thermo Fischer Scientific, 81 Wyman Street, Waltham, MA 02454 USA).

(2) DNA library preparation

DNA libraries were prepared starting from 500 ng input RNA using an FGFR Fusion Detection kit for Illumina^{R} assembled by ArcherDx (now Enzymatics Inc., Beverly, MA). This detection kit is similar to the Archer^{TM} ALK, RET, ROS1 Fusion Detection v1 for Illumina^{R} Platform marketed by Enzymatics (product no. AK0001-8) except that ALK-, RET- and ROS1-specific primers are replaced with primers specific for human FGFR1, FGFR2 and FGFR3. The ArcherDx fusion detection kits use anchored multiplex PCR (AMP^{TM}) and temperature-stable reagents in order to create libraries for targeted sequencing on the Illumina^{R} MiSeq instrument (Illumina, 5200 Illumina Way, San Diego, CA 92122 USA). Libraries were prepared following the manufacturer's instructions.

The concentration of each bar-coded library was determined by PCR using Kapa Biosystems Library Quantification Kit for Illumina no. KK4824 (Kapa Biosystems, Inc., Wilmington, MA) according to the manufacturer's instructions. Bar-coded libraries were pooled at equimolar concentrations, loaded on an IlluminaR MiSeq desktop sequencer at 10 pM each and sequenced using the Illumina MiSeq v2 (300 cycles) reagent kit (MS-102-2002, Illumina Inc., San Diego, CA) and Nextera workflow chemistry. 15% PhiX

control v3 of Illumina (FC-110-3001) was added at 10pM to the library pool to serve as a sequencing control.

(3) Analysis of sequencing results

Sequence reads were stripped of the adapter sequences at the 3'-end using cutadapt (Martin, M. (2011) EMBnet.journal 17: 10-12). Read sequences exceeding 20 nucleotides in length were then mapped onto the human genome sequence (Genome Reference Consortium Human Build 37 (GRCh37) using bowtie (Langmead. B. et al. (2009). Genome Biology 10: R25) and tophat (Trapnell, C. et al. (2009) Bioinformatics 25: 1105-11). Reads mapping to more than two locations were discarded and the remainder were annotated to Ensembl genes (Flicek, P. et al. (2014) Nucleic Acids. Res. 42; D749-55, Database issue).

In a second step, not-mapped reads exceeding 75 nucleotides in length were split into three parts of equal length, and left and right extreme parts were mapped separately onto the genome using the same method as above, and then annotated to Ensembl genes.

Fusion sequences were identified as sequences with both reads (from paired-end sequencing) mapped onto two different genes (one being a FGFR gene) or with not-mapped read(s) annotated to two different genes after split (one being a FGFR gene). The consensus sequence of each fusion was generated from multiple alignment of all reads from fusion sequences with Clustal W (Larkin, M.A. et al. (2007) *Bioinformatics*, 23, 2947-2948). (See SEQ ID NOs: 1 and 2.)

All data processing steps were performed using R 3.0.2 (http://www.r-project.org/) and Bioconductor packages (Gentleman, R.C. et al. (2004) Genome Biology 5: R80).

The fusion gene sequences in SEQ ID NOs: 1 and 2 and the derived polypeptide sequences in SEQ ID NOs: 3 and 4 resulted from the above-described effort. Based on the latter polynucleotide sequences, PCR primers were designed that were used in the characterization experiments described under Example 2.

Example 2: Characterization of fusion genes of the invention

RNA samples (700 ng each) (prepared from two tumor biopsies as described in section (1) of Example 1) were denatured at 65°C for 5 min and then reverse-transcribed with random hexamer primers in a final volume of 20µL using the Roche Transcriptor First Strand cDNA synthesis kit (product no. 04 896 866 001; Roche Diagnostics AG, Rotkreuz, Switzerland). Reverse transcription was performed with the following cycling conditions: 25°C for 10 min, 55°C for 30 min and 85°C for 5 min.

PCR amplifications were carried out in a 20µL reaction volume containing 2 µL of ten-fold diluted cDNA, 0.3 µM of each forward and reverse primer (Table 1) and Roche FastStart PCR Master 1x (product no. 04 710 444 001; Roche Diagnostics AG). Cycling conditions were as follows: 1 cycle at 95°C for 4 min followed by 40 cycles of 95°C for 30 s, 50°C for 30 s and 72°C for 1 min, and 1 cycle at 72°C for 7 min. PCR products were separated on Lonza FlashGel DNA Cassette 2.2% agarose gels (product no. 57031; Lonza Ltd., Basel, Switzerland).

Table 1: Forward and reverse primers used in PCR amplifications. Primers were designed based on the nucleotide sequences of SEQ ID NOs: 1-2 using Primer3 software (Steve Rozen and Helen J. Skaletsky (2000) *Primer3 on the WWW for general users and for biologist programmers.* In: Krawetz S, Misener S (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp 365-386)

Code	SEQ ID NO.	Primer Name	Sequence (5' to 3')
primer 16	8	FGFR2_Forward	CAGAGACCAACGTTCAAGCA
primer 17	9	FGFR2_Reverse1	GGTTGGCTGAGGTCCAAGTA
primer 18	10	FGFR2_Reverse2	TCTTGTGTCAGGGTAACTAGGTGA
primer 19	11	VCL_Reverse_1	AGCTTGATTTCCAGGGTTCC
primer 20	12	VCL_Reverse_2	AAGATACGAGCAGCCGAGAC
primer 21	14	CCDC147_Reverse1	TTCCAGGACTTGCTTTCCAC
primer 22	15	CCDC147_Reverse2	TCCCTGAAAATCTCTTTCCATT

The results of this analysis can be gleaned from the pre-stained gels shown in Fig. 1. Amplification of cDNA prepared from one of the two tumor biopsies yielded FGFR2-VCL amplification products with the predicted lengths of about 89 and 114 bp, respectively,

depending on which of the two reverse primers was used for PCR. No FGFR2-CCDC147 amplification product was detected. Amplification of cDNA prepared from the other tumor biopsy resulted in FGFR2- CCDC147 amplification products with the predicted lengths of about 93 and 132 bp, respectively, depending on which of two reverse primers was used for PCR. No FGFR2-VCL amplification product was detected. Aliquots of amplification products were purified using QIAgen MinElute PCR Purification Kit (28004) according to the manufacturer's instructions and were sequenced using the Sanger method (Sanger et al. (1975) J. Mol. Biol. 94: 441–8). Results are represented in Table 2.

Table 2: Nucleotide sequence analysis of PCR amplification products. The 3' most nucleotides of the FGFR2-coding sequence are in bold and underlined.

Sequencing Primer	Fusion	SEQ ID NO	Sequencing results (5' to 3')
16	FGFR2-VCL	73	CAACCAATGA**GG**TGGTCTCGGCTGCTCGTATCTTACTTAGGAACCCTGGAAATCAAGCTGA
19	FGFR2-VCL	74	TTGGTTGTGAGAGTGAGAATTCGATCCAAGTCTTCTACCAACTGCTTGAACGTTGGTCTCTGAAGNC
16	FGFR2-CCDC147	75	CCAATGA**GG**AAAAGGGTGGAAAGCAAGTCCTGGAAGAATCTGCATTTGAAGAAATGGAAAGAGATTTTCAGGGAAACT
22	FGFR2-CCDC147	76	GTGAGAGTGAGANTTCGATCCAAGTCTTCTACCAACTGCTTGAACGTTGGTCTCTGAAGT
16	native FGFR2	77	AATGA**GG**AATACTTGGACCTCAGCCAACCTCTCGAACAGTATTCACCTAGTTACCCTGACACAAGAAGAAN
18	native FGFR2	78	TTCCTCATTGGTTGTGAGAGTGAGAATTCGATCCAAGTCTTCTACCAACTGCTTGAACGTTGGTCTCTGAACA

Larger PCR amplification products (of about 400 and 300 bp, respectively) were obtained for the FGFR2-VCL fusion gene transcript using primer pairs 42a/20 and 16/29 (See Tables 1 and 3). These products were sequenced using the Sanger method. The nucleotide sequence determined is shown as SEQ ID NO: 5; the derived polypeptide fragment sequence is SEQ ID NO: 6.

Table 3: Additional forward and reverse primers used in PCR amplifications

Code	SEQ ID NO	Primer Name	Sequence (5' to 3')
primer 42a	7	FGFR2 forward	GACTTTGGACTCGCCAGAGA
primer 29	13	VCL_Reverse	CCAGCAACCAGCATCTGAG

Example 3: Diagnostic method for detecting FGFR2-VCL and FGFR2-CCDC147 fusions in tumor biopsies

(1) Biopsy samples and RNA preparation

Ten-micron slides are first prepared from formalin-fixed, paraffin-embedded clinical specimens of solid tumors using methods well known in the field. After hematoxylin and eosin staining, a tumoral portion of the tissue is macro-dissected and subjected to total RNA extraction using the High Pure FFPE RNA isolation kit (Roche, catalog number #06 650 775 001), following the manufacturer's instructions. RNA quantity is assessed using a NanoDrop spectrophotometer (Thermo Fischer Scientific, 81 Wyman Street, Waltham, MA 02454 USA).

(2) Sequencing

FGFR2-targeted cDNA libraries composed of 100-300bp sequences are prepared from 500 ng of total RNA using an ArcherDx NGS library preparation kit (Enzymatics, Suite 407J, 100 Cummings Center, Beverly, MA 01910, USA), following the manufacturer's instructions. Basically, primers specific to wild type FGFR2 are used to select RNA sequences containing corresponding sequences. The libraries are subjected to paired-end sequencing of 50-150-bp fragments using an Illumina^{R} MiSeq instrument (Illumina, 5200 Illumina Way, San Diego, CA 92122 USA) as instructed by the manufacturer.

(3) Detection of FGFR2-VCL and FGFR2-CCDC147 fusion genes

Reads obtained are aligned with known subsequences of FGFR2 transcripts beginning at the 3' end of the tyrosine kinase domain and ending at the 3' end of the transcripts. Such subsequences are provided in SEQ ID NOs: 64-67 and 72. Sequence alignment software such as BLAST can be used. Alignment length should be equal to or greater than 18. In case an alignment obtained is with the complementary strand, the complementary sequence of the reads should be considered for further analysis instead of the initial reads.

Reads not matching the FGFR2 subsequences should be discarded. In a case in which no read matches any FGFR2 subsequence, fusion detection should be considered to be inconclusive.

For each alignment with an FGFR2 subsequence, the nucleotide position of the read corresponding to the 3' end of the alignment is named N. The read subsequence corresponding to positions N+1 to N+18 is then extracted. In the case that the read does not contain this subsequence, the alignment is discarded. The 18-base subsequence corresponding to positions N+1 to N+18 is then aligned with subsequences of VCL and CCDC147 transcripts beginning at the start codon and ending at the beginning of the sequences encoding the most distant oligomerization domain. Such subsequences are provided in SEQ ID NOs: 68-71.

A tumor specimen is considered to be positive for an FGFR2-VCL or an FGFR2-CCDC147 fusion gene if a read is identified that has no more than 5 mismatches, preferably no more than 3-4 mismatches, more preferably no more than 1-2 mismatches and most preferably no mismatches over the 18-base stretch corresponding to positions N+1 to N+18 with subsequences of VCL or CCDC147 transcripts, respectively. If no such alignment is observed for any read, the tumor specimen is considered to be negative for FGFR2-VCL and FGFR2-CCDC147 fusion genes.

Example 4: In vitro tumorigenicity

(1) Establishment of cell pools stably expressing fusion polypeptides

Stable Rat2 cell pools stably expressing FGFR2-CCDC147 or FGFR2-VCL fusions, respectively, were generated using pExoIN2-based expression plasmids ExoIN2-FGFR2_CCDC147 and pExoIN2-FGFR2_VCL. pExoIN2 was obtained from Trenzyme, (Germany). The latter expression plasmids were introduced into Rat2 cells by electroporation (LONZA Nucleofector II Device / program [X-005], Solution R). 24h post-transduction, cells were subjected to 1.5µg/mL puromycin to derive stable expressor cell pools. The FGFR2-CCDC147 sequence used in this example was composed of SEQ ID NO: 16 for the FGFR2 part (nucleic acid 1 to nucleic acid 2574) and SEQ ID NO: 20 for the CCDC147 part (nucleic acid 156 to nucleic acid 3313). The FGFR2-VCL sequence used was composed of SEQ ID NO: 16 for the FGFR2 part (nucleic acid 1 to nucleic acid 2574) and SEQ ID NO: 18 for the VCL part (nucleic acid 2117 to nucleic acid 5482).

(2) Assessment of anchorage-independent growth properties of stable expressor cell pools

Single cell suspensions were prepared using Accutase (GE Healthcare Europe GmbH) and diluted such that appropriate cell numbers were seeded in 6well dishes in 0.4% soft-agar top layer without selection antibiotic (seeding densities (cells/well): 10.000, 3.000, 1.000, 300, 100 and 30). Dishes were incubated in a 5% CO_2 environment at 37°C for colony formation. After 21 days of incubation, colonies were fixed using 10% (v/v) acetic acid and 10% (v/v) methanol in H_2O and stained with crystal violet (0,01% (w/v) in H_2O). Plating efficiency was determined as a ratio between the number of colonies observed after 21 days of incubation in soft agar and the number of seeded cells. The plating efficiency results are shown in Figure 2.

For FGFR2-VCL expressing cells, strong colony formation was observed exhibiting high plating efficiency (approximately 50%). Plating efficiency of FGFR2-CCDC147 expressing cells was below that of parental Rat2 cells but colony sizes of parental Rat2 cell line were smaller compared to those of FGFR2-VCL and FGFR2-CCDC147 expressing cell lines (Figure 3).

Example 5: In vitro sensitivity to FGFR inhibitors

(1) Cell proliferation assay using FACS

24h after seeding (25000 cells/wells), Rat2 cells obtained as per Example 4 (either parental cells or cells expressing FGFR2-VCL or FGFR2-CCDC147 fusion polypeptide) FGFR inhibitors were added, and the cultures were incubated for another 72h in the presence of the inhibitors. At the end of the incubation period, cells were counted by FACS. IC50 values were calculated in Graphpad Prism 6 using sigmoidal response (variable slope) curve fit. The inhibitors used were selective FGFR inhibitors Compound A, BGJ398 and AZD4547, as well as multi-kinase inhibitor Ponatinib.

FACS analysis revealed that proliferation of the fusion polypeptide-expressing cells was inhibited by all FGFR inhibitors tested (Figure 4). Multi-kinase inhibitor Ponatinib also inhibited cell proliferation of parental cells due to its large and non-specific spectrum of activity. Relative IC50s of >1024 nM, 227.1nM and 14.7nM were obtained for Compound A in Rat2 parental cells, FGFR2-VCL expressing cells and FGFR2-CCDC147 expressing cells, respectively.

Hence, the above results show that cells expressing either fusion of the present invention (FGFR2-VCL or FGFR2-CCDC147, respectively) are sensitive to FGFR selective inhibitors in vitro, whereas parental cells are only sensitive to a multi-kinase inhibitor.

(2) Cell proliferation assay using Cell Titer Glo

Rat2 cells obtained as per Example 4 (either parental cells or cells expressing FGFR2-VCL or FGFR2-CCDC147 fusion polypeptide) were used in this assay. Cells were seeded in 96 well plates and cultured for 24h before addition of FGFR inhibitors (the same as in section (1) above). After 72h of further incubation, cell growth was analyzed by determination of the cellular ATP content (Cell Titer Glo; Promega) using a luminescence plate reader. Relative IC50s of >3000 nM, 0.53nM and 73.4nM were obtained for Compound A in Rat2 parental cells, FGFR2-VCL expressing cells and FGFR2-CCDC147 expressing cells, respectively. Dose response curves and a summary of IC50s are shown in Figure 5.

Thus, FGFR selective inhibitors showed potent inhibition of proliferation of both cell lines harboring FGFR2 fusions (FGFR2-VCL and FGFR2-CCDC147, respectively), whereas parental cells were not affected by FGFR selective inhibitors.

Example 6: In vivo tumorigenicity

In vivo tumorigenicity of Rat2 cells (either parental cells or cells expressing FGFR2-VCL or FGFR2-CCDC147 fusion polypeptide, obtained as per Example 4) was evaluated in vivo in a subcutaneous tumor model in 5-6 week-old female NMRI nude mice. The study consisted of 6 experimental groups, each containing 5 animals. $5x10^6$ and $1x10^6$ Rat2 parental cells (Groups 1 and 2), Rat2-FGFR2-CCDC147 cells (Groups 3 and 4) and Rat2-FGFR2-VCL cells (Groups 5 and 6), respectively, were subcutaneously implanted on day 0. Animal weights of all groups increased continuously during the course of the study. Primary tumor volumes were determined twice weekly by caliper measurement. Tumor volumes were calculated according to the formula $W^2xL/2$ (L= length and W= the perpendicular width of the tumor, L>W). Results are shown on Figure 6 (Data are displayed as means ± SEM).

In the case of Rat2 parental cells (control), no primary tumor growth could be observed, regardless of the inoculum size. In the case of Rat2-FGFR2-CCDC147 cells, substantial

tumor growth could be observed starting around day 28, and the animals that had received higher numbers of implanted cells ($5x10^6$) exhibited faster tumor growth. In the case of Rat2-FGFR2-VCL cells (Groups 5 and 6), tumor growth was observed starting around 10 days after implantation. Owing to fast tumor growth, Group 5 ($5x10^6$ cells) had to be terminated for ethical reasons (tumor burden) on day 21, and Group 6 ($1x10^6$ cells) on day 28. Tumors expressing either FGFR2 fusion construct (FGFR2-VCL or FGFR2-CCDC147) were therefore shown to be tumorigenic in vivo in female NMRI nude mice.

Example 7: In vivo sensitivity to FGFR inhibitors

(1) FGFR2-CCDC147 fusion

On Day 0, $5x10^6$ FGFR2-CCDC147-expressing Rat2 cells in 100µl PBS were subcutaneously implanted into the left flank of 5-6 week-old female NMRI nude mice (group size of 6 animals). Compound A was orally administered once daily for 14 consecutive days (30 or 60 mg/kg) in mice with established tumors (D25, mean Tumor Volume = 135 mm^3). Primary tumor sizes were measured twice weekly by calipering. Animals were terminated on last day of treatment and tumors weighed at necropsy.

Compound A showed potent antitumor efficacy in vivo in the FGFR2-CCDC147 expressing model (Figure 7A). Compound A inhibited tumor growth at the two tested doses (30 and 60mg/kg), whereas no significant effect was observed on body weight (panel B). Data are displayed as means ± SEM. P-values were calculated compared to the Vehicle Control using the Mann Whitney test (unpaired t-test in parentheses).

Thus, FGFR selective inhibitor Compound A, administered daily for 14 consecutive days at 30 and 60 mg/kg, showed a highly significant antitumoral efficacy (comparable at both doses) in the subcutaneously implanted Rat2-FGFR2-CCDC147 tumor model in female NMRI nude mice in vivo.

(2) FGFR2-VCL fusion

On Day 0, $1x10^6$ FGFR2-VCL expressing Rat2 cells in 100µl PBS were subcutaneously implanted into the left flank of 5-6 week-old female NMRI nude mice (group size of 6 animals). Compound A was orally administered once daily for 14 consecutive days (30 or 60 mg/kg) in mice with established tumors (D15, mean Tumor Volume = 188 mm^3).

Primary tumor sizes were measured twice weekly by calipering. Animals were terminated on last day of treatment and tumors weighed at necropsy.

FGFR selective inhibitor Compound A showed potent antitumor efficacy in vivo in FGFR2-VCL-expressing model (Figure 8A). Compound A inhibited tumor growth without significantly affecting body weight (panel B). Data are displayed as means ± SEM. P-values were calculated compared to the Vehicle Control and between Groups 2 and 3 using the Mann Whitney test (unpaired t-test in parentheses).

Thus, FGFR selective inhibitor Compound A, administered orally daily for 14 consecutive days at 30 and 60 mg/kg, showed a highly significant and dose-dependent antitumoral efficacy in the subcutaneously implanted Rat2-FGFR2-VCL tumor model in female NMRI nude mice in vivo. Tumor stasis was observed for the 60 mg/kg-treated group.

Example 8: Functional analysis of FGFR2-VCL / FGFR2-CCDC147 autophosphorylation

In this example, the following sequences were used: the FGFR2-CCDC147 fusion gene was composed of SEQ ID NO: 16 for the FGFR2 part (nucleic acid 1 to nucleic acid 2574) and SEQ ID NO: 20 for the CCDC147 part (nucleic acid 156 to nucleic acid 3313); and the FGFR2-VCL fusion gene was composed of SEQ ID NO: 16 for the FGFR2 part (nucleic acid 1 to nucleic acid 2574) and SEQ ID NO: 18 for the VCL part (nucleic acid 2117 to nucleic acid 5482).

(1) Autophosphorylation of FGFR2 fusion polypeptides in transiently transfected HEK293T cells assessed by Western blot

HEK293T cells were transiently transfected with expression plasmids containing either an FGFR2-VCL or an FGFR2-CCDC147 fusion gene, which genes had been supplemented with a C-terminal double myc-tag extension. As a negative control, HEK293T cells were mock-transfected. Cells were treated for 90 min with 0.1% DMSO, 1E-05 M Vargatef or 1E-05 M Compound A. After treatment, expression and autophosphorylation of the fusion polypeptides was analysed by Western blotting. The expression of myc-tagged fusion polypeptide was determined using the anti-myc antibody 9E10 (α-myc), and the phosphorylation analysis was performed with the anti-phospho-tyrosine antibody pY99 (α-pY). The results are shown in Figure 9. Both FGFR2-VCL- and FGFR2-CCDC147-expressing cells show ligand-independent high

levels of FGFR phosphorylation. This autophosphorylation appears substantially reduced in FGFR2-VCL-expressing cells and, somewhat more marginally, in FGFR2-CCDC147-expressing cells upon exposure to FGFR selective inhibitor Compound A or less selective inhibitor Vargatef.

(2) Autophosphorylation of FGFR2 fusion polypeptides in transiently transfected HEK293T cells measured by ELISA

HEK293T cells, mock-transfected (control) or transiently transfected with expressible FGFR2-VCL or FGFR2-CCDC147 fusion genes, were treated with 0.1% DMSO, 1E-05 M Vargatef or 1E-05 M Compound A for 90 min. After treatment, autophosphorylation of the FGFR2 fusion polypeptides was analysed using a sandwich phosphotyrosine ELISA. Each condition was done in duplicates. Mean Optical Densities (OD) values of each condition are presented in Figure 10.

Both FGFR2-VCL- and FGFR2-CCDC147-expressing cells show ligand-independent high levels of FGFR phosphorylation, which levels are decreased by FGFR selective inhibitor Compound A or less selective inhibitor Vargatef.

Example 9: Diagnostic method for detecting FGFR2-VCL and FGFR2-CCDC147 fusions in tumor biopsies using digital detection of mRNA

Total RNA was extracted from FFPE samples of human cholangiocarcinoma as described in Example 1(1). 500 ng of total RNA were used per sample and analyzed using the nCounter Gene Expression Assay protocol as instructed by the manufacturer (www.nanostring.com). The nCounter assay is based on direct digital detection of mRNA molecules of interest using target-specific, color-coded probes that hybridize directly to a target molecule in solution, so that the expression level of each gene is measured in a relative fashion by counts, without the need for cDNA synthesis and amplification. Each probe is constituted by a Reporter probe part of 50 bases that carries the barcode and a Capture probe part of 50 bases that carries a biotin molecule allowing the target/probe complex to be immobilized to a streptavidin-coated nCounter Cartridge for data collection (Counts) after washout of excess probes.

For each fusion gene of the present invention, probes were designed and synthetized by NanoString (Custom CodeSet), then inserted with all consumables and reagents in a ready-to-use nCounter Master Kit for sample processing in the nCounter Analysis

System. The target sequences used for the design of the probes, as provided by NanoString, are indicated in Table 4 (detail of Capture probe and Reporter probe not known).

Table 4: Target sequences used for the design of the probes in the nCounter assay.

Gene Identifier	Target Sequence
FGFR2- CCDC147	AGCAGTTGGTAGAAGACTTGGATCGAATTCTCACTCTCACAAC CAATGAGGAAAAGGGTGGAAAGCAAGTCCTGGAAGAATCTGC ATTTGAAGAAATGGA
FGFR2- VCL	AGCAGTTGGTAGAAGACTTGGATCGAATTCTCACTCTCACAAC CAATGAGGTGGTCTCGGCTGCTCGTATCTTACTTAGGAACCC TGGAAATCAAGCTGC

Count values obtained were background corrected and normalized against the most stable housekeeping genes as previously described (Beaume, et al. (2011) J Microbiol Methods 84: 327–334). The results are presented in Table 5.

Table 5: Normalized values obtained for the 2 fusion genes of the present invention in intrahepatic cholangiocarcinoma (iCCA) samples

Sample Identifier	Fusion name	
	FGFR2-VCL	FGFR2-CCDC147
iCCA FFPE sample #1	129	4652
iCCA FFPE sample #2	66204	1

It could thus be determined that the iCCA FFPE sample #1 harbors a FGFR2-CCDC147 gene fusion and that the iCCA FFPE sample #2 harbors a FGFR2-VCL gene fusion. For iCCA FFPE sample #1, it was determined (by PCR amplification using validated primers for both fusions) that the normalized value of 129 regarding FGFR2-VCL resulted from nonspecific hybridization and constituted background noise.

Example 10: Diagnostic method for detecting FGFR2-VCL and FGFR2-CCDC147 fusions in tumor biopsies using DNA

(1) Biopsy samples and DNA preparation

Total DNA are extracted from macro-dissected 10 µm thick sections from formalin-fixed, paraffin-embedded tissue obtained from biopsies of human cholangiocarcinomas. DNA extraction is performed using an FFPE DNA Isolation Kit.

(2) Sequencing

Fusions are detected by capture-enriched DNA sequencing using capture probes for? the VCL and CCDC147 genes, as described by Duncavage et al. (Duncavage et al. Mod Pathol. 2012 Jun;25(6):795-804). After fragmentation of the genomic DNA (to fragments of about 250 to 500 bp in length), the fragmented DNA is end repaired, ligated to adapters as per manufacturer's protocol (Illumina, San Diego, CA, USA). Sequencing libraries are then hybridized with the capture probes per manufacturer's instructions (Agencourt Bioscience, Beverly, MA, USA). The enriched DNA is then amplified using universal primers targeting the adapters. DNA is then subjected to paired-end sequencing of 50-150-bp fragments using an Illumina^{R} MiSeq instrument (Illumina, 5200 Illumina Way, San Diego, CA 92122 USA) as instructed by the manufacturer.

(3) Detection of FGFR2-VCL and FGFR2-CCDC147 fusion genes

The detection of the FGFR2-VCL and FGFR2-CCDC147 fusion genes is performed as described in example 3(3).

CLAIMS

1. A cDNA encoding a fusion polypeptide comprising an FGFR2 polypeptide and a CCDC147 polypeptide or a VCL polypeptide, wherein the FGFR2 polypeptide is the whole or a part of a wildtype FGFR2 polypeptide, or the whole or a part of a mutant FGFR2 polypeptide having one or more amino acid substitutions, deletions, or insertions with respect to the wildtype polypeptide; the CCDC147 polypeptide is the whole or a part of a wildtype CCDC147 polypeptide, or the whole or a part of a mutant CCDC147 polypeptide having one or more amino acid substitutions, deletions, or insertions with respect to the wildtype polypeptide; and the VCL polypeptide is the whole or a part of a wildtype VCL polypeptide, or the whole or a part of a mutant VCL polypeptide having one or more amino acid substitutions, deletions, or insertions with respect to the wildtype polypeptide.

2. A cDNA encoding a fusion polypeptide comprising an FGFR2 polypeptide and a CCDC147 polypeptide or a VCL polypeptide, wherein the cDNA combines a first polynucleotide encoding the whole or a part of a wildtype FGFR2 polypeptide or a mutant polynucleotide derived therefrom by substitution, deletion or insertion of one or more codons; and a second polynucleotide encoding the whole or a part of a wildtype CCDC147 polypeptide or VCL polypeptide or a mutant polynucleotide derived therefrom by substitution, deletion or insertion of one or more codons.

3. The cDNA of claim 1 or 2, wherein the FGFR2 polypeptide includes a complete tyrosine kinase domain.

4. The cDNA of claim 2, wherein the first polynucleotide contains FGFR2 exons 1-16 and part or all of FGFR2 exon 17, and the second polynucleotide lacks CCDC147 exon 1 in the case of an FGFR2-CCDC147 fusion or lacks VCL exons 1-14 in the case of an FGFR2-VCL fusion.

5. The cDNA of any one of claims 1-4, wherein the cDNA is derived from a gene transcript isolated from a human cholangiocarcinoma.

6. A vector carrying the cDNA of any of claims 1-5.

7. The vector of claim 6, wherein the vector is an expression vector.

8. A prokaryotic or eukaryotic cell transduced with the vector of claim 6 or claim 7.

9. The cell of claim 8, wherein the cell is an *E. coli* cell.

10. The cell of claim 8, wherein the cell is a mammalian cell.

11. The polynucleotide sequence of SEQ ID NO: 1.

12. The polynucleotide sequence of SEQ ID NO: 2.

13. A fusion polypeptide comprising an FGFR2 polypeptide and a CCDC147 polypeptide or a VCL polypeptide, wherein the FGFR2 polypeptide is the whole or a part of a wildtype FGFR2 polypeptide, or the whole or a part of a mutant FGFR2 polypeptide with one or more amino acid substitutions, deletions, or insertions in the wildtype polypeptide; the CCDC147 polypeptide is the whole or a part of a wildtype CCDC147 polypeptide, or the whole or a part of a mutant CCDC147 polypeptide with one or more amino acid substitutions, deletions, or insertions in the wildtype polypeptide; and the VCL polypeptide is the whole or a part of a wildtype VCL polypeptide, or a part of a mutant VCL polypeptide with one or more amino acid substitutions, deletions, or insertions in the wildtype polypeptide, wherein the fusion polypeptide is a recombinant polypeptide, is isolated from a cancer cell that is propagated *in vitro* or as a xenograft, or is isolated from a human cholangiocarcinoma.

14. An antibody or antigen-binding fragment that binds the fusion polypeptide of claim 1.

15. A primer pair consisting of a sense and an antisense primer, which primers are capable of specifically hybridizing to the cDNA of any one of claims 1-5 and of directing replication of the cDNA or parts thereof.

16. An oligonucleotide probe capable of specifically hybridizing to the cDNA of any one of claims 1-5, to a gene for the fusion polypeptide of claim 1 or to a transcript of this gene.

17. An antisense oligonucleotide capable of hybridization in a life cell to a messenger RNA encoding the fusion polypeptide of claim 1, which hybridization has the effect of preventing or reducing translation of the messenger RNA.

18. A siRNA directed to a messenger RNA encoding the fusion polypeptide of claim 1.

19. A kit for detecting a fusion polypeptide according to claim 1, comprising one or more antibodies or antigen-binding fragments capable of binding the fusion polypeptide.

20. A kit for detecting a gene or a transcript of a gene encoding a fusion polypeptide according to claim 1, comprising a primer pair according to claim 15 or an oligonucleotide probe according to claim 16.

21. An inhibitor of FGFR kinase activity for use in a therapeutic regimen in a subject suffering from cancer, wherein said subject contains or expresses a gene encoding a fusion polypeptide according to claim 1 or expresses a fusion polypeptide according to claim 1.

22. The inhibitor of claim 21, wherein the inhibitor of FGFR kinase activity is selected from the group consisting of 5-amino-1-(2-methyl-1H-benzimidazol-5-yl)-1H-pyrazol-4-yl]-(1H-indol-2-yl)-methanone, PD173074, Pazopanib, AZD4547, Ponatinib, Dovitinib, BGJ398, E-3810, JNJ-42756493, ARQ 087, LY2874455, BAY1163877, ASP5878, E7090, ODM-203, Nintedanib, TAS-120, PRN 1109 and PRN 1371.

23. The inhibitor of claim 21, wherein the inhibitor of FGFR kinase activity is 5-amino-1-(2-methyl-1H-benzimidazol-5-yl)-1H-pyrazol-4-yl]-(1H-indol-2-yl)-methanone.

24. A method of personalized cancer therapy, comprising subjecting a subject containing or expressing a gene for a fusion polypeptide according to claim 1 or expressing a fusion polypeptide according to claim 1 to a therapeutic regimen that comprises administration of a pharmaceutical composition comprising (1) an inhibitor of FGFR kinase activity, (2) an antibody or antigen-binding fragment according to claim 14, (3) an antisense oligonucleotide according to claim 17 or (4) an siRNA according to claim 18.

25. A method of personalized cancer therapy, comprising the steps of

(a) taking a biopsy or fluid sample containing cancer cells or tumor circulating DNA from a subject suffering from cancer;

(b) determining whether the cells in the biopsy or fluid sample contain or express a gene encoding a fusion polypeptide according to claim 1 or express a fusion polypeptide according to claim 1;

(c) selecting the subject containing or expressing the gene for the fusion polypeptide or expressing the fusion polypeptide for the treatment of step d; and

(d) subjecting the selected subject to a therapeutic regimen that comprises administration of a pharmaceutical composition comprising (1) an inhibitor of FGFR kinase activity, (2) an antibody or antigen-binding fragment according to claim 14, (3) an antisense oligonucleotide according to claim 17 or (4) a siRNA according to claim 18.

26. The method of claim 24 or 25, wherein the inhibitor of FGFR kinase activity is selected from the group consisting of 5-amino-1-(2-methyl-1H-benzimidazol-5-yl)-1H-pyrazol-4-yl]-(1H-indol-2-yl)-methanone, PD173074, Pazopanib, AZD4547, Ponatinib, Dovitinib, BGJ398, E-3810, JNJ-42756493, ARQ 087, LY2874455, BAY1163877, ASP5878, E7090, ODM-203, Nintedanib, TAS-120, PRN 1109 and PRN 1371.

27. The method of claim 24 or 25, wherein the inhibitor of FGFR kinase activity is 5-amino-1-(2-methyl-1H-benzimidazol-5-yl)-1H-pyrazol-4-yl]-(1H-indol-2-yl)-methanone.

28. A method of characterizing a tumor in a human subject suffering from cancer, comprising assaying protein or nucleic acids of a biopsy or fluid sample containing cancer cells or tumor circulating DNA from said subject to identify the presence or absence of a gene encoding a fusion polypeptide according to claim 1 or an expressed fusion polypeptide according to claim 1.

29. The method according to claim 28, wherein said cancer is cholangiocarcinoma.

30. A method for identifying a compound having FGFR inhibitory activity comprising the steps of:

(a) culturing a cell that expresses a fusion polypeptide according to claim 1 and whose growth is dependent on this expression in the presence or absence of a test compound and determining the level of cell proliferation;

(b) comparing the proliferation level of the cultured cell in the presence and absence of the test compound; and

(c) judging that the test compound has FGFR inhibitory activity when the proliferation level of the cell cultured in the presence of the test compound is lower than that of the cell cultured in the absence of the test compound.

31. The method of claim 30, wherein the cell is a cancer cell.

32. The method of claim 30, wherein the cancer cell is a cholangiocarcinoma cell.