Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2880013 A1 2014/01/30

(21) 2 880 013

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2013/07/24
(87) Date publication PCT/PCT Publication Date: 2014/01/30
(85) Entrée phase nationale/National Entry: 2015/01/23
(86) N° demande PCT/PCT Application No.: US 2013/051888
(87) N° publication PCT/PCT Publication No.: 2014/018673
(30) Priorité/Priority: 2012/07/24 (US61/675,006)

(51) Cl.Int./Int.Cl. *C12N 15/62* (2006.01),
A61K 31/167 (2006.01), *A61K 31/404* (2006.01),
A61K 31/416 (2006.01), *A61K 31/4439* (2006.01),
A61K 31/47 (2006.01), *A61K 31/4709* (2006.01),
A61K 31/496 (2006.01), *A61K 31/5025* (2006.01),
A61K 31/505 (2006.01), *A61K 31/506* (2006.01),
A61K 31/517 (2006.01), ...

(71) Demandeur/Applicant:
THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK, US

(72) Inventeurs/Inventors:
IAVARONE, ANTONIO, US;
LASORELLA, ANNA, US; ...

(54) Titre : PROTEINES DE FUSION ET LEURS PROCEDES ASSOCIES
(54) Title: FUSION PROTEINS AND METHODS THEREOF

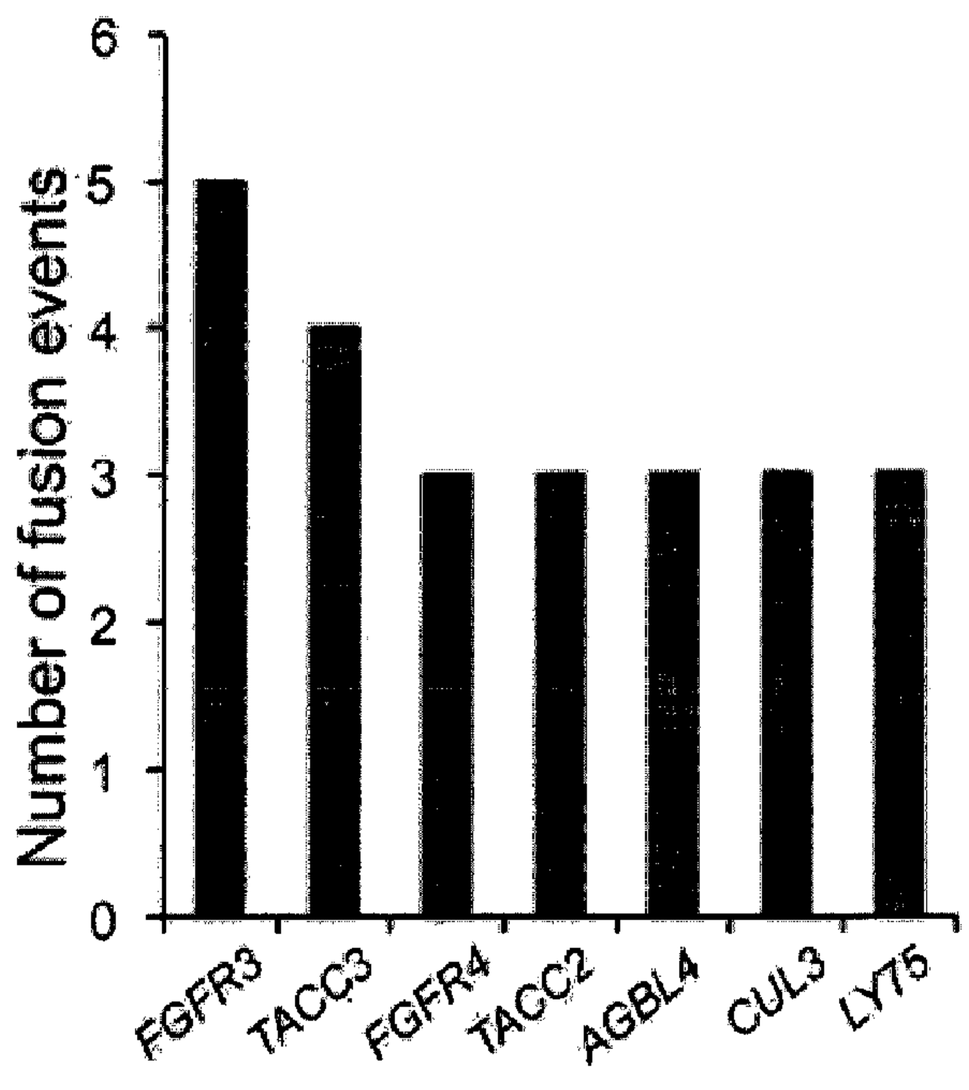


FIG. 1A

(57) Abrégé/Abstract:
The invention discloses oncogenic fusion proteins. The invention provides methods for treating gene-fusion based cancers.

(21) 2 880 013

(13) A1

(51) Cl.Int./Int.Cl. (suite/continued) *A61K 31/519* (2006.01), *A61K 31/53* (2006.01), *A61K 39/395* (2006.01), *A61P 35/00* (2006.01), *C07H 21/00* (2006.01), *C07K 14/705* (2006.01), *C07K 14/71* (2006.01), *C07K 16/28* (2006.01), *C07K 16/40* (2006.01), *C07K 19/00* (2006.01), *C12N 9/12* (2006.01), *C12Q 1/68* (2006.01), *G01N 33/574* (2006.01)

(72) Inventeurs(suite)/Inventors(continued): RABANDAN, RAUL, US

(74) Agent: GOWLING LAFLEUR HENDERSON LLP

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 January 2014 (30.01.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/018673 A3

(51) International Patent Classification:
G01N 33/574 (2006.01) *A61P 35/00* (2006.01)
G01N 33/68 (2006.01) *C07H 21/00* (2006.01)

(21) International Application Number:
PCT/US2013/051888

(22) International Filing Date:
24 July 2013 (24.07.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/675,006 24 July 2012 (24.07.2012) US

(71) Applicant: **THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK** [US/US]; 412 Low Memorial Library, New York, NY 10027 (US).

(72) Inventors: **IAVARONE, Antonio**; 245 East 72nd Street, Apt. 14a, New York, NY 10021 (US). **LASORELLA, Anna**; 245 East 72nd Street, Apt. 14a, New York, NY 10021 (US). **RABANDAN, Raul**; 1130 St. Nicholas Avenue, New York, NY 10027 (US).

(74) Agents: **LOVE, Jane, M.** et al.; Wilmer Cutler Pickering Hale And Dorr LLP, 7 World Trade Center, 250 Greenwich Street, New York, NY 10007 (US).

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

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

[Continued on next page]

(54) Title: FUSION PROTEINS AND METHODS THEREOF

(57) Abstract: The invention discloses oncogenic fusion proteins. The invention provides methods for treating gene-fusion based cancers.

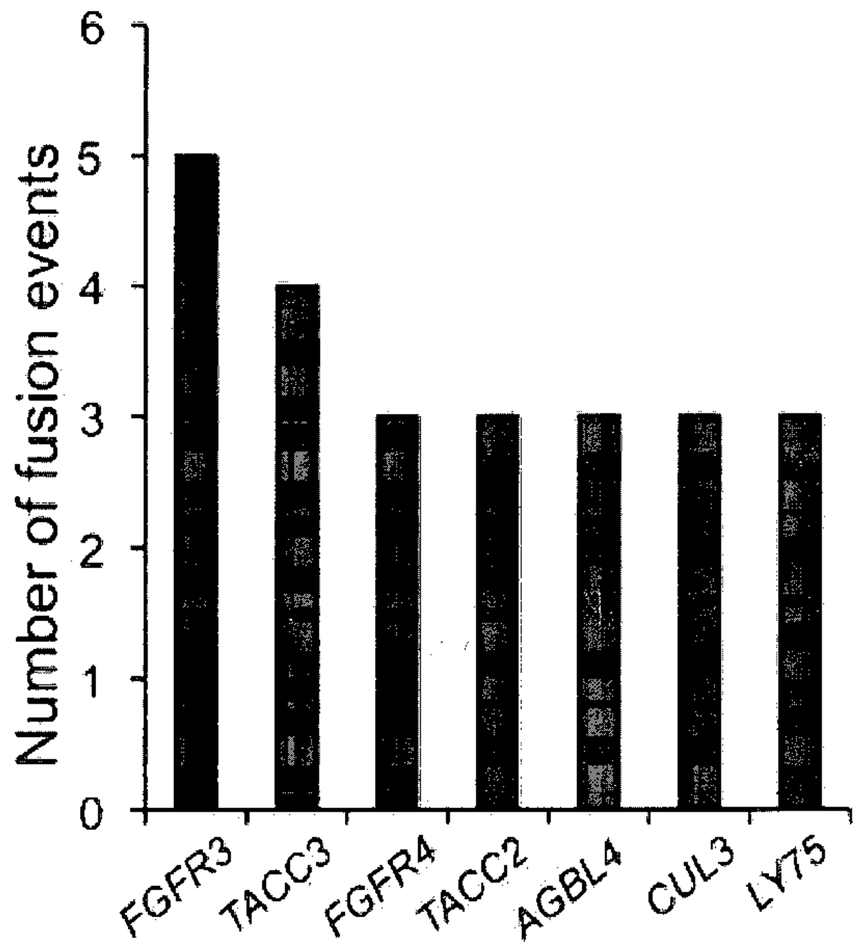


FIG. 1A

WO 2014/018673 A3

WO 2014/018673 A3

Published:

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

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

(88) Date of publication of the international search report:
17 April 2014

FUSION PROTEINS AND METHODS THEREOF

[0001] This application claims the benefit of and priority to U.S. Provisional Patent Application No. 61/675,006, filed on July 24, 2012, the content of which is hereby incorporated by reference in its entirety.

[0002] All patents, patent applications and publications cited herein are hereby incorporated by reference in their entirety. The disclosures of these publications in their entireties are hereby incorporated by reference into this application.

[0003] This patent disclosure contains material that is subject to copyright protection. The copyright owner has no objection to the facsimile reproduction by anyone of the patent document or the patent disclosure as it appears in the U.S. Patent and Trademark Office patent file or records, but otherwise reserves any and all copyright rights.

GOVERNMENT SUPPORT

[0004] This invention was made with government support under Grant Nos. R01CA101644 and R01CA085628 awarded by the National Cancer Institute. The Government has certain rights in the invention.

BACKGROUND OF THE INVENTION

[0005] Glioblastoma multiforme (GBM) is the most common form of brain cancer and among the most incurable and lethal of all human cancers. The current standard of care includes surgery, chemotherapy, and radiation therapy. However, the prognosis of GBM remains uniformly poor. There are few available targeted therapies and none that specifically target GBM.

[0006] The target population of GBM patients who may carry FGFR-TACC gene fusions and would benefit from targeted inhibition of FGFR kinase activity is estimated to correspond to 6,000 patients per year world-wide.

SUMMARY OF THE INVENTION

[0007] The invention is based, at least in part, on the discovery of a highly expressed class of gene fusions in GBM, which join the tyrosine kinase domain of FGFR genes to the TACC domain of TACC1 or TACC3. The invention is based, at least in part, on the finding

that FGFR-TACC fusions identify a subset of GBM patients who will benefit from targeted inhibition of the tyrosine kinase activity of FGFR. Identification of fusions of FGFR and TACC genes in glioblastoma patients and other subjects afflicted with a gene-fusion associated cancer (such as an epithelial cancer) are useful therapeutic targets.

[0008] An aspect of the invention provides for a purified fusion protein comprising a tyrosine kinase domain of an FGFR protein fused to a polypeptide that constitutively activates the tyrosine kinase domain of the FGFR protein. In one embodiment, the FGFR protein is FGFR1, FGFR2, FGFR3, or FGR4. In another embodiment, the purified fusion protein is essentially free of other human proteins.

[0009] An aspect of the invention provides for a purified fusion protein comprising a transforming acidic coiled-coil (TACC) domain fused to a polypeptide with a tyrosine kinase domain, wherein the TACC domain constitutively activates the tyrosine kinase domain. In one embodiment, the TACC protein is TACC1, TACC2, or TACC3. In another embodiment, the purified fusion protein is essentially free of other human proteins.

[0010] An aspect of the invention provides for a purified fusion protein comprising the tyrosine kinase domain of an FGFR protein fused 5' to the TACC domain of a transforming acidic coiled-coil-containing (TACC) protein. In one embodiment, the FGFR protein is FGFR1, FGFR2, FGFR3, or FGR4. In another embodiment, the TACC protein is TACC1, TACC2, or TACC3. In another embodiment, the purified fusion protein is essentially free of other human proteins.

[0011] An aspect of the invention provides for a purified fusion protein encoded by an FGFR1-TACC1 nucleic acid, wherein FGFR1-TACC1 comprises a combination of exons 1-17 of FGFR1 located on human chromosome 8p11 spliced 5' to a combination of exons 7-13 of TACC1 located on human chromosome 8p11, wherein a genomic breakpoint occurs in any one of exons 1-17 of FGFR1 and any one of exons 7-13 of TACC1. In another embodiment, the purified fusion protein is essentially free of other human proteins.

[0012] An aspect of the invention provides for a purified fusion protein encoded by an FGFR2-TACC2 nucleic acid, wherein FGFR2-TACC2 comprises a combination of any exons 1-18 of FGFR2 located on human chromosome 10q26 spliced 5' to a combination of any exons 1-23 of TACC2 located on human chromosome 10q26. In another embodiment, the purified fusion protein is essentially free of other human proteins.

[0013] An aspect of the invention provides for a purified fusion protein encoded by an FGFR3-TACC3 nucleic acid, wherein FGFR3-TACC3 comprises a combination of exons 1-16 of FGFR3 located on human chromosome 4p16 spliced 5' to a combination of exons 8-16 of TACC3 located on human chromosome 4p16, wherein a genomic breakpoint occurs in any one of exons 1-16 of FGFR3 and any one of exons 8-16 of TACC3. In another embodiment, the purified fusion protein is essentially free of other human proteins.

[0014] An aspect of the invention provides for a purified fusion protein encoded by an FGFR3-TACC3 nucleic acid, wherein FGFR3-TACC3 comprises a combination of exons 1-16 of FGFR3 located on human chromosome 4p16 spliced 5' to a combination of exons 8-16 of TACC3 located on human chromosome 4p16, wherein a genomic breakpoint occurs in any one of ***introns*** 1-16 of FGFR3 and any one of exons 8-16 of TACC3. In another embodiment, the purified fusion protein is essentially free of other human proteins.

[0015] An aspect of the invention provides for a purified fusion protein encoded by an FGFR3-TACC3 nucleic acid, wherein FGFR3-TACC3 comprises a combination of exons 1-16 of FGFR3 located on human chromosome 4p16 spliced 5' to a combination of exons 8-16 of TACC3 located on human chromosome 4p16, wherein a genomic breakpoint occurs in any one of exons 1-16 of FGFR3 and any one of ***introns*** 7-16 of TACC3. In another embodiment, the purified fusion protein is essentially free of other human proteins.

[0016] An aspect of the invention provides for a purified fusion protein encoded by an FGFR3-TACC3 nucleic acid, wherein FGFR3-TACC3 comprises a combination of exons 1-16 of FGFR3 located on human chromosome 4p16 spliced 5' to a combination of exons 8-16 of TACC3 located on human chromosome 4p16, wherein a genomic breakpoint occurs in any one of ***introns*** 1-16 of FGFR3 and any one of ***introns*** 7-16 of TACC3. In another embodiment, the purified fusion protein is essentially free of other human proteins.

[0017] An aspect of the invention provides for a synthetic nucleic acid encoding the fusion proteins described above.

[0018] An aspect of the invention provides for a purified FGFR3-TACC3 fusion protein comprising SEQ ID NO: 79, 158, 159, 160, or 161. In another embodiment, the purified fusion protein is essentially free of other human proteins.

[0019] An aspect of the invention provides for a purified FGFR3-TACC3 fusion protein

having a genomic breakpoint comprising at least 3 consecutive amino acids from amino acids 730-758 of SEQ ID NO: 90 and comprising at least 3 consecutive amino acids from amino acids 549-838 of SEQ ID NO: 92. In another embodiment, the purified fusion protein is essentially free of other human proteins.

[0020] An aspect of the invention provides for a purified FGFR3-TACC3 fusion protein having a genomic breakpoint comprising SEQ ID NO: 78. In another embodiment, the purified fusion protein is essentially free of other human proteins.

[0021] An aspect of the invention provides for a purified FGFR3-TACC3 fusion protein having a genomic breakpoint comprising any one of SEQ ID NOS: 85, 86, 87, or 89. In another embodiment, the purified fusion protein is essentially free of other human proteins.

[0022] An aspect of the invention provides for a purified FGFR1-TACC1 fusion protein comprising SEQ ID NO: 150. In another embodiment, the purified fusion protein is essentially free of other human proteins.

[0023] An aspect of the invention provides for a purified FGFR1-TACC1 fusion protein having a genomic breakpoint comprising at least 3 consecutive amino acids from amino acids 746-762 of SEQ ID NO: 146 and comprising at least 3 consecutive amino acids from amino acids 572-590 of SEQ ID NO: 148. In another embodiment, the purified fusion protein is essentially free of other human proteins.

[0024] An aspect of the invention provides for a purified FGFR1-TACC1 fusion protein having a genomic breakpoint comprising SEQ ID NO: 88. In another embodiment, the purified fusion protein is essentially free of other human proteins.

[0025] An aspect of the invention provides for a purified DNA encoding an FGFR3-TACC3 fusion protein comprising SEQ ID NO: 94. In another embodiment, the purified fusion protein is essentially free of other human proteins.

[0026] An aspect of the invention provides for a synthetic nucleic acid encoding an FGFR3-TACC3 fusion protein having a genomic breakpoint comprising at least 9 consecutive in-frame nucleotides from nucleotides 2443-2530 of SEQ ID NO: 91 and comprising at least 9 consecutive in-frame nucleotides from nucleotides 1800-2847 of SEQ ID NO: 93.

[0027] An aspect of the invention provides for a synthetic nucleic acid encoding an FGFR3-TACC3 fusion protein having a genomic breakpoint comprising any one of SEQ ID NOS: 1-77.

[0028] An aspect of the invention provides for a synthetic nucleic acid encoding an FGFR1-TACC1 fusion protein comprising SEQ ID NO: 151.

[0029] An aspect of the invention provides for a synthetic nucleic acid encoding an FGFR1-TACC1 fusion protein having a genomic breakpoint comprising at least 9 consecutive in-frame nucleotides from nucleotides 3178-3228 of SEQ ID NO: 147 and comprising at least 9 consecutive in-frame nucleotides from nucleotides 2092-2794 of SEQ ID NO: 149.

[0030] An aspect of the invention provides for a synthetic nucleic acid encoding an FGFR1-TACC1 fusion protein having a genomic breakpoint comprising SEQ ID NO: 83.

[0031] An aspect of the invention provides for an antibody or antigen-binding fragment thereof, that specifically binds to a purified fusion protein comprising a tyrosine kinase domain of an FGFR protein fused to a polypeptide that constitutively activates the tyrosine kinase domain of the FGFR protein. In one embodiment, the FGFR protein is FGFR1, FGFR2, FGFR3, or FGFR4. In another embodiment, the fusion protein is an FGFR-TACC fusion protein. In a further embodiment, the FGFR-TACC fusion protein is FGFR1-TACC1, FGFR2-TACC2, or FGFR3-TACC3. In some embodiments, the FGFR1-TACC1 fusion protein comprises the amino acid sequence of SEQ ID NO: 150. In other embodiments, the FGFR3-TACC3 fusion protein comprises the amino acid sequence of SEQ ID NO: 79, 158, 159, 160, or 161.

[0032] An aspect of the invention provides for a composition for decreasing in a subject the expression level or activity of a fusion protein comprising the tyrosine kinase domain of an FGFR protein fused to a polypeptide that constitutively activates the tyrosine kinase domain of the FGFR protein, the composition in an admixture of a pharmaceutically acceptable carrier comprising an inhibitor of the fusion protein. In one embodiment, the fusion protein is an FGFR-TACC fusion protein. In another embodiment, the inhibitor comprises an antibody that specifically binds to a FGFR-TACC fusion protein or a fragment thereof; a small molecule that specifically binds to a FGFR protein; a small molecule that specifically binds to a TACC protein; an antisense RNA or antisense DNA that decreases

expression of a FGFR-TACC fusion polypeptide; a siRNA that specifically targets a FGFR-TACC fusion gene; or a combination of the listed inhibitors. In a further embodiment, the FGFR protein is FGFR1, FGFR2, FGFR3, or FGFR4. In some embodiments, the FGFR-TACC fusion protein is FGFR1-TACC1, FGFR2-TACC2, or FGFR3-TACC3. In other embodidments, the small molecule that specifically binds to a FGFR protein comprises AZD4547, NVP-BGJ398, PD173074, NF449, TK1258, BIBF-1120, BMS-582664, AZD-2171, TSU68, AB1010, AP24534, E-7080, LY2874455, or a combination of the listed small molecules.

[0033] An aspect of the invention provides for a method for decreasing in a subject in need thereof the expression level or activity of a fusion protein comprising the tyrosine kinase domain of an FGFR protein fused to a polypeptide that constitutively activates the tyrosine kinase domain of the FGFR protein. In one embodiment, the method comprises administering to the subject a therapeutic amount of a composition for decreasing the expression level or activity in a subject of a fusion protein comprising the tyrosine kinase domain of an FGFR protein fused to a polypeptide that constitutively activates the tyrosine kinase domain of the FGFR protein. In one embodiment, the method comprises obtaining a sample from the subject to determine the level of expression of an FGFR fusion molecule in the subject. In some embodiments, the sample is incubated with an agent that binds to an FGFR fusion molecule, such as an antibody, a probe, a nucleic acid primer, and the like. In one embodiment, the detection or determining comprises nucleic acid sequencing, selective hybridization, selective amplification, gene expression analysis, or a combination thereof. In another embodiment, the detection or determination comprises protein expression analysis, for example by western blot analysis, ELISA, or other antibody detection methods. In a further embodiment, the method comprises determining whether the fusion protein expression level or activity is decreased compared to fusion protein expression level or activity prior to administration of the composition, thereby decreasing the expression level or activity of the fusion protein. In one embodiment, the fusion protein is an FGFR-TACC fusion protein. In a further embodiment, the FGFR protein is FGFR1, FGFR2, FGFR3, or FGFR4. In some embodiments, the FGFR-TACC fusion protein is FGFR1-TACC1, FGFR2-TACC2, or FGFR3-TACC3. In one embodiment, the composition for decreasing the expression level or activity of a fusion protein comprises an antibody that specifically binds to a FGFR-TACC fusion protein or a fragment thereof; a small molecule that specifically binds to a FGFR protein; a small molecule that specifically binds to a TACC protein; an

antisense RNA or antisense DNA that decreases expression of a FGFR-TACC fusion polypeptide; a siRNA that specifically targets a FGFR-TACC fusion gene; or a combination of the listed inhibitors. In a further embodiment, the FGFR protein is FGFR1, FGFR2, FGFR3, or FGFR4. In some embodiments, the FGFR-TACC fusion protein is FGFR1-TACC1, FGFR2-TACC2, or FGFR3-TACC3. In other embodiments, the small molecule that specifically binds to a FGFR protein comprises AZD4547, NVP-BGJ398, PD173074, NF449, TK1258, BIBF-1120, BMS-582664, AZD-2171, TSU68, AB1010, AP24534, E-7080, LY2874455, or a combination of the small molecules listed.

[0034] An aspect of the invention provides for a method for treating a gene-fusion associated cancer in a subject in need thereof, the method comprising administering to the subject an effective amount of a FGFR fusion molecule inhibitor. In one embodiment, the gene-fusion associated cancer comprises an epithelial cancer. In one embodiment, the gene-fusion associated cancer comprises glioblastoma multiforme, breast cancer, lung cancer, prostate cancer, or colorectal carcinoma. In one embodiment, the method comprises obtaining a sample from the subject to determine the level of expression of an FGFR fusion molecule in the subject. In some embodiments, the sample is incubated with an agent that binds to an FGFR fusion molecule, such as an antibody, a probe, a nucleic acid primer, and the like. In one embodiment, the detection or determining comprises nucleic acid sequencing, selective hybridization, selective amplification, gene expression analysis, or a combination thereof. In another embodiment, the detection or determination comprises protein expression analysis, for example by western blot analysis, ELISA, or other antibody detection methods. In another embodiment, the FGFR fusion protein comprises an FGFR protein fused to a polypeptide that constitutively activates the tyrosine kinase domain of the FGFR protein. In one embodiment, the fusion protein is an FGFR-TACC fusion protein. In another embodiment, the inhibitor comprises an antibody that specifically binds to a FGFR-TACC fusion protein or a fragment thereof; a small molecule that specifically binds to a FGFR protein; a small molecule that specifically binds to a TACC protein; an antisense RNA or

antisense DNA that decreases expression of a FGFR-TACC fusion polypeptide; a siRNA that specifically targets a FGFR-TACC fusion gene; or a combination of the listed inhibitors. In a further embodiment, the FGFR protein is FGFR1, FGFR2, FGFR3, or FGFR4. In some embodiments, the FGFR-TACC fusion protein is FGFR1-TACC1, FGFR2-TACC2, or FGFR3-TACC3. In other embodiments, the small molecule that specifically binds to a FGFR protein comprises AZD4547, NVP-BGJ398, PD173074, NF449, TK1258, BIBF-1120, BMS-582664, AZD-2171, TSU68, AB1010, AP24534, E-7080, LY2874455, or a combination of the small molecules listed.

[0035] An aspect of the invention provides for a method of decreasing growth of a solid tumor in a subject in need thereof, the method comprising administering to the subject an effective amount of a FGFR fusion molecule inhibitor, wherein the inhibitor decreases the size of the solid tumor. In one embodiment, the solid tumor comprises glioblastoma multiforme, breast cancer, lung cancer, prostate cancer, or colorectal carcinoma. In one embodiment, the method comprises obtaining a sample from the subject to determine the level of expression of an FGFR fusion molecule in the subject. In some embodiments, the sample is incubated with an agent that binds to an FGFR fusion molecule, such as an antibody, a probe, a nucleic acid primer, and the like. In one embodiment, the detection or determining comprises nucleic acid sequencing, selective hybridization, selective amplification, gene expression analysis, or a combination thereof. In another embodiment, the detection or determination comprises protein expression analysis, for example by western blot analysis, ELISA, or other antibody detection methods. In another embodiment, the FGFR fusion protein comprises an FGFR protein fused to a polypeptide that constitutively activates the tyrosine kinase domain of the FGFR protein. In one embodiment, the fusion protein is an FGFR-TACC fusion protein. In another embodiment, the inhibitor comprises an antibody that specifically binds to a FGFR-TACC fusion protein or a fragment thereof; a small molecule that specifically binds to a FGFR protein; a small molecule that specifically binds to a TACC protein; an antisense RNA or antisense DNA that decreases expression of a FGFR-TACC fusion polypeptide; a siRNA that specifically targets a FGFR-TACC fusion gene; or a combination of the listed inhibitors. In a further embodiment, the FGFR protein is FGFR1, FGFR2, FGFR3, or FGFR4. In some embodiments, the FGFR-TACC fusion protein is FGFR1-TACC1, FGFR2-TACC2, or FGFR3-TACC3. In other embodiments, the

small molecule that specifically binds to a FGFR protein comprises AZD4547, NVP-BGJ398, PD173074, NF449, TK1258, BIBF-1120, BMS-582664, AZD-2171, TSU68, AB1010, AP24534, E-7080, LY2874455, or a combination of the small molecules listed.

[0036] An aspect of the invention provides for a diagnostic kit for determining whether a sample from a subject exhibits a presence of a FGFR fusion, the kit comprising at least one oligonucleotide that specifically hybridizes to a FGFR fusion, or a portion thereof. In one embodiment, the oligonucleotides comprise a set of nucleic acid primers or *in situ* hybridization probes. In another embodiment, the oligonucleotide comprises SEQ ID NO: 162, 163, 164,165, 166, 167, 168, 169, or a combination of the listed oligonucleotides. In one embodiment, the primers prime a polymerase reaction only when a FGFR fusion is present. In another embodiment, the determining comprises gene sequencing, selective hybridization, selective amplification, gene expression analysis, or a combination thereof. In a further embodiment, the FGFR-fusion is an FGFR-TACC fusion. In some embodiments, the FGFR is FGFR1, FGFR2, FGFR3, or FGFR4. In other embodiments, the FGFR-TACC fusion is FGFR1-TACC1, FGFR2-TACC2, or FGFR3-TACC3.

[0037] An aspect of the invention provides for a diagnostic kit for determining whether a sample from a subject exhibits a presence of a FGFR fusion protein, the kit comprising an antibody that specifically binds to a FGFR fusion protein comprising SEQ ID NO: 79, 85, 86, 87, 88, 89, 150, 158, 159, 160, or 161, wherein the antibody will recognize the protein only when a FGFR fusion protein is present. In one embodiment, the antibody directed to and FGFR fusion comprising SEQ ID NO: 79, 85, 86, 87, 88, 89, 150, 158, 159, 160, or 161. In a further embodiment, the FGFR-fusion is an FGFR-TACC fusion. In some embodiments, the FGFR is FGFR1, FGFR2, FGFR3, or FGFR4. In other embodiments, the FGFR-TACC fusion is FGFR1-TACC1, FGFR2-TACC2, or FGFR3-TACC3.

[0038] An aspect of the invention provides for a method for detecting the presence of a FGFR fusion in a human subject. In one embodiment, the method comprises obtaining a biological sample from the human subject. In some embodiments, the sample is incubated with an agent that binds to an FGFR fusion molecule, such as an antibody. In another embodiment, the detection or determination comprises protein expression analysis, for example by western blot analysis, ELISA, or other antibody detection methods. In some embodiments, the method further comprises assessing whether to administer a FGFR fusion molecule inhibitor based on the expression pattern of the subject. In further embodiments,

the method comprises administering a FGFR fusion molecule inhibitor to the subject. In another embodiment, the method comprises detecting whether or not there is a FGFR fusion present in the subject. In one embodiment, the detecting comprises measuring FGFR fusion protein levels by ELISA using an antibody directed to SEQ ID NO: 79, 85, 86, 87, 88, 89, 150, 158, 159, 160, or 161; western blot using an antibody directed to SEQ ID NO: 79, 85, 86, 87, 88, 89, 150, 158, 159, 160, or 161; mass spectroscopy, isoelectric focusing, or a combination of the listed methods. In some embodiments, the FGFR-fusion is an FGFR-TACC fusion. In other embodiments, the FGFR is FGFR1, FGFR2, FGFR3, or FGFR4. In other embodiments, the FGFR-TACC fusion is FGFR1-TACC1, FGFR2-TACC2, or FGFR3-TACC3.

[0039] An aspect of the invention provides for a method for detecting the presence of a FGFR fusion in a human subject. In one embodiment, the method comprises obtaining a biological sample from a human subject. In some embodiments, the sample is incubated with an agent that binds to an FGFR fusion molecule, such as a probe, a nucleic acid primer, and the like. In other embodiments, the detection or determination comprises nucleic acid sequencing, selective hybridization, selective amplification, gene expression analysis, or a combination thereof. In some embodiments, the method further comprises assessing whether to administer a FGFR fusion molecule inhibitor based on the expression pattern of the subject. In further embodiments, the method comprises administering a FGFR fusion molecule inhibitor to the subject. In another embodiment, the method comprises detecting whether or not there is a nucleic acid sequence encoding a FGFR fusion protein in the subject. In one embodiment, the nucleic acid sequence comprises any one of SEQ ID NOS: 1-77, 80-84, or 95-145. In another embodiment, the detecting comprises using hybridization, amplification, or sequencing techniques to detect a FGFR fusion. In a further embodiment, the amplification uses primers comprising SEQ ID NO: 162, 163, 164,165, 166, 167, 168, or 169. In some embodiments, the FGFR-fusion is an FGFR-TACC fusion. In other embodiments, the FGFR is FGFR1, FGFR2, FGFR3, or FGFR4. In other embodiments, the FGFR-TACC fusion is FGFR1-TACC1, FGFR2-TACC2, or FGFR3-TACC3.

[0040] An aspect of the invention provides for a method of initiating oncogenic transformation *in vitro*. The method comprises (a) transducing cells cultured *in vitro* with FGFR-TACC fusion DNA; and (b) determining whether the cells acquire the ability to grow in anchorage-independent conditions, form multi-layered foci, or a combination thereof.

[0041] An aspect of the invention provides for a method of initiating oncogenic transformation *in vivo*. The method comprises (a) transducing cells cultured *in vitro* with FGFR-TACC fusion DNA; (b) injecting a mouse with the transduced cells; and (c) determining whether a tumor grows in the mouse. In one embodiment, the injecting is a subcutaneous or intracranial injection.

[0042] An aspect of the invention provides a method of identifying a compound that decreases the oncogenic activity of a FGFR-TACC fusion. The method comprises (a) transducing a cell cultured *in vitro* with FGFR-TACC DNA; (b) contacting a cell with a ligand source for an effective period of time; and (c) determining whether the cells acquire the ability to grow in anchorage-independent conditions, form multi-layered foci, or a combination thereof, compared to cells cultured in the absence of the test compound.

BRIEF DESCRIPTION OF THE FIGURES

[0043] To conform to the requirements for PCT patent applications, many of the figures presented herein are black and white representations of images originally created in color. In the below descriptions and the examples, the colored plots, and images are described in terms of its appearance in black and white. The original color versions can be viewed in Singh *et al.*, *Science* (2012), 337(6099):1231-5 (including the accompanying Supplementary Information available in the on-line version of the manuscript available on the Science web site). For the purposes of the PCT, the contents of Singh *et al.*, *Science* (2012), 337(6099):1231-5, including the accompanying "Supplementary Information," are herein incorporated by reference.

[0044] **FIG. 1A** is a graph that shows genes recurrently involved in gene fusions in TCGA. Only genes involved in at least three gene fusions across different samples are displayed.

[0045] **FIG. 1B** shows an *FGFR3-TACC3* gene fusion identified by whole transcriptome sequencing of GSCs. 76 split-reads (SEQ ID NOS: 2-77, respectively) are shown aligning on the breakpoint. The predicted reading frame at the breakpoint is shown at the top (FGFR3 nucleotide sequence in purple (left) and TACC3 nucleotide sequence in green (right); SEQ ID NO: 1) with FGFR3 sequences below the predicted reading frame in red (left) and TACC3 in blue (right). The putative amino acid sequence (SEQ ID NO: 78) corresponding to SEQ ID NO: 1 is shown above the predicted reading frame.

[0046] **FIG. 1C** shows an *FGFR3-TACC3* gene fusion identified by whole transcriptome sequencing of GSCs. On the left, *FGFR3-TACC3*-specific PCR from cDNA derived from GSCs and GBM is shown. On the right, Sanger sequencing chromatogram shows the reading frame at the breakpoint (SEQ ID NO: 80) and putative translation of the fusion protein (SEQ ID NO: 85) in the positive samples.

[0047] **FIG. 1D** shows an *FGFR3-TACC3* gene fusion identified by whole transcriptome sequencing of GSCs. Amino acid sequence of the FGFR3-TACC3 protein is shown (SEQ ID NO: 79). Residues corresponding to FGFR3 or to TACC3 are shown in green or red (underlined), respectively. The fusion protein joins the tyrosine kinase domain of FGFR3 to the TACC domain of TACC3.

[0048] **FIG. 1E** shows an *FGFR3-TACC3* gene fusion identified by whole transcriptome sequencing of GSCs. Genomic fusion of *FGFR3* exon 17 with intron 7 of *TACC3* is shown. In the fused mRNA, exon 16 of *FGFR3* is spliced 5' to exon 8 of *TACC3*. Filled arrows indicate the position of the fusion-genome primers, which generate fusion-specific PCR products in GSC-1123 and GBM-1123.

[0049] **FIG. 2A** shows recurrent gene fusions between *FGFR* and *TACC* genes in GBM. Specifically, *FGFR3-TACC3* gene fusions are shown that were identified by exome sequencing analysis. Split-reads are shown aligning the genomic breakpoints of *FGFR3* and *TACC3* genes in the four TCGA GBM samples. For TCGA-27-1835, SEQ ID NO: 95 shows the reading frame at the breakpoint (bold), while SEQ ID NOS: 96-107, respectively, show alignments of the genomic breakpoints. For TCGA-19-5958, SEQ ID NO: 108 shows the reading frame at the breakpoint (bold), while SEQ ID NOS: 109-111, respectively, show alignments of the genomic breakpoints. For TCGA-06-6390, SEQ ID NO: 112 shows the reading frame at the breakpoint (bold), while SEQ ID NOS: 113-131, respectively, show alignments of the genomic breakpoints. For TCGA-12-0826, SEQ ID NO: 132 shows the reading frame at the breakpoint (bold), while SEQ ID NOS: 133-145, respectively, show alignments of the genomic breakpoints.

[0050] **FIG. 2B** shows recurrent gene fusions between *FGFR* and *TACC* genes in GBM. On the left, a gel of *FGFR-TACC*-specific PCR is shown for *FGFR3-TACC3* from a GBM cDNA sample. On the right, Sanger sequencing chromatograms show the reading frame at the breakpoint (SEQ ID NO: 81) and putative translation of the fusion protein (SEQ ID NO: 86) in the positive samples.

[0051] **FIG. 2C** shows recurrent gene fusions between *FGFR* and *TACC* genes in GBM. On the left, a gel of *FGFR-TACC*-specific PCR is shown for *FGFR3-TACC3* from a GBM cDNA sample. On the right, Sanger sequencing chromatograms show the reading frame at the breakpoint (SEQ ID NO: 82) and putative translation of the fusion protein (SEQ ID NO: 87) in the positive samples.

[0052] **FIG. 2D** shows recurrent gene fusions between *FGFR* and *TACC* genes in GBM. Co-outlier expression of *FGFR3* and *TACC3* in four GBM tumors from Atlas-TCGA is shown in the plot.

[0053] **FIG. 2E** shows recurrent gene fusions between *FGFR* and *TACC* genes in GBM. CNV analysis shows micro-amplifications of the rearranged portions of the *FGFR3* and *TACC3* genes in the same four Atlas-TCGA GBM samples.

[0054] **FIG. 2F** shows recurrent gene fusions between *FGFR* and *TACC* genes in GBM. On the left, a gel of *FGFR-TACC*-specific PCR is shown for *FGFR1-TACC1* from a GBM cDNA sample. On the right, Sanger sequencing chromatograms show the reading frame at the breakpoint (SEQ ID NO: 83) and putative translation of the fusion protein (SEQ ID NO: 88) in the positive samples.

[0055] **FIG. 2G** shows recurrent gene fusions between *FGFR* and *TACC* genes in GBM. On the left, a gel of *FGFR-TACC*-specific PCR is shown for *FGFR3-TACC3* from a GBM cDNA sample. On the right, Sanger sequencing chromatograms show the reading frame at the breakpoint (SEQ ID NO: 84) and putative translation of the fusion protein (SEQ ID NO: 89) in the positive samples.

[0056] **FIG. 3A** shows transforming activity of FGFR-TACC fusion proteins. FGFR1-TACC1 and FGFR3-TACC3 induce anchorage-independent growth in Rat1A fibroblasts. The number of soft agar colonies was scored from triplicate samples 14 days after plating. Representative microphotographs are shown.

[0057] **FIG. 3B** are photomicrographs showing of immunofluoresence staining of tumors from mice injected with *Ink4A;Arf-/-* astrocytes expressing FGFR3-TACC3 showing positivity for glioma-specific (Nestin, Olig2 and GFAP) and proliferation markers (Ki67 and pHH3). Sub-cutaneous tumors were generated by *Ink4A;Arf-/-* astrocytes expressing FGFR-TACC fusions.

[0058] **FIG. 3C** shows Kaplan-Meier survival curves of mice injected intracranially with pTomo-shp53 (n = 8) or pTomo-EGFRvIII-shp53 (n = 7) (green line; "light grey" in black

and white image) and pTomo-FGFR3-TACC3-shp53 (n = 8, red line; "dark grey" in black and white image). Points on the curves indicate deaths (log-rank test, p = 0.00001, pTomo-shp53 vs. pTomo-FGFR3-TACC3-shp53).

[0059] **FIG. 3D** shows representative photomicrographs of Hematoxylin and Eosin staining of advanced FGFR3-TACC3-shp53 generated tumors showing histological features of high-grade glioma. Of note is the high degree of infiltration of the normal brain by the tumor cells. Immunofluorescence staining shows that glioma and stem cell markers (Nestin, Olig2 and GFAP), the proliferation markers (Ki67 and pHH3) and the FGFR3-TACC3 protein are widely expressed in the FGFR3-TACC3-shp53 brain tumors. F1-T1: FGFR1-TACC1; F3-T3: FGFR3-TACC3; F3-T3-K508M: FGFR3-TACC3-K508M.

[0060] **FIG. 4A** shows that FGFR3-TACC3 localizes to spindle poles, delays mitotic progression and induces chromosome segregation defects and aneuploidy Constitutive auto-phosphorylation of FGFR3-TACC3 fusion. *Ink4A;Arf-/-* astrocytes transduced with empty lentivirus or a lentivirus expressing FGFR3-TACC3 or FGFR3-TACC3-K508M were left untreated (0) or treated with 100 nM of the FGFR inhibitor PD173074 for the indicated times. Phospho-proteins and total proteins were analyzed by Western blot using the indicated antibodies.

[0061] **FIG. 4B** shows that FGFR3-TACC3 localizes to spindle poles, delays mitotic progression and induces chromosome segregation defects. Photomicrographs are shown of confocal microscopy analysis of FGFR3-TACC3 in *Ink4A;Arf-/-* astrocytes. Maximun intensity projection of z-stacked images shows FGFR3-TACC3 (red; "dark grey" in black and white image) coating the spindle poles of a representative mitotic cell (upper panels). In telophase (lower panels) FGFR3-TACC3 localizes to the mid-body. α-tubulin (green; "grey" in black and white image), DNA (DAPI, blue; "light grey" in black and white image).

[0062] **FIG. 4C** shows representative fluorescence video-microscopy for cells transduced with vector or FGFR3-TACC3.

[0063] **FIG. 4D** shows a Box and Whisker plot representing the analysis of the time from nuclear envelope breakdown (NEB) to anaphase onset and from NEB to nuclear envelope reconstitution (NER). The duration of mitosis was measured by following 50 mitoses for each condition by time-lapse microscopy.

[0064] **FIG. 4E** shows that FGFR3-TACC3 localizes to spindle poles, delays mitotic progression and induces chromosome segregation defects. Representative images are shown

of cells with chromosome missegregation. Arrows point to chromosome misalignments, lagging chromosomes, and chromosome bridges.

[0065] **FIG. 4F** shows quantitative analysis of segregation defects in Rat1A expressing FGFR1-TACC1 and FGFR3-TACC3. F3-T3: FGFR3-TACC3; F3-T3-K508M: FGFR3-TACC3-K508M.

[0066] **FIG. 5A** shows karyotype analysis of Rat1A cells transduced with control, FGFR3, TACC3 or FGFR3-TACC3 expressing lentivirus. Distribution of chromosome counts of cells arrested in mitosis and analyzed for karyotypes using DAPI. Chromosomes were counted in 100 metaphase cells for each condition to determine the ploidy and the diversity of chromosome counts within the cell population. FGFR3-TACC3 fusion induces aneuploidy.

[0067] **FIG. 5B** shows representative karyotypes and **FIG. 5C** shows distribution of chromosome counts of human astrocytes transduced with control or FGFR3-TACC3 expressing lentivirus. Chromosomes were counted in 100 metaphase cells for each condition to determine the ploidy and the diversity of chromosome counts within the cell population.

[0068] **FIG. 5D** shows quantitative analysis of chromosome number in 100 metaphase cells for each condition to determine the ploidy and the diversity of chromosome counts within the cell population. (n=3 independent experiments).

[0069] **FIG. 6A** shows inhibition of FGFR-TK activity corrects the aneuploidy initiated by FGFR3-TACC3. The upper panel is a karyotype analysis of Rat1A cells transduced with control or FGFR3-TACC3 lentivirus and treated with vehicle (DMSO) or PD173470 (100 nM) for five days. The lower panel shows the ploidy and the diversity of chromosome counts within the cell population were determined by quantitative analysis of chromosome number in 100 metaphase cells for each condition.

[0070] **FIG. 6B** shows inhibition of FGFR-TK activity corrects the aneuploidy initiated by FGFR3-TACC3. Correction of premature sister chromatid separation (PMSCS) by PD173470 in cells expressing FGFR3-TACC3. Panels show representative metaphase spreads. DNA was stained by DAPI. **FIG. 6C** shows quantitative analysis of metaphases with loss of sister chromatid cohesion (n=3; $p = 0.001$, FGFR3-TACC3 treated with DMSO vs. FGFR3-TACC3 treated with PD173470).

[0071] **FIG. 7A** shows inhibition of FGFR-TK activity suppresses tumor growth initiated by FGFR3-TACC3. Growth rate of Rat1A transduced with the indicated lentiviruses and

treated for three days with increasing concentrations of PD173074. Cell growth was determined by the MTT assay. Data are presented as the means±standard error (n=4).

[0072] **FIG. 7B** shows the growth rate of GSC-1123 treated with PD173470 at the indicated concentrations for the indicated times. Cell growth was determined by the MTT assay. Data are presented as the means±standard error (n=4).

[0073] **FIG. 7C** shows the growth inhibitory effect of silencing FGFR3-TACC3 fusion. At the left, parallel cultures of GSC-1123 cells were transduced in triplicate. Rat1A cells expressing FGFR3-TACC3 fusion were transduced with lentivirus expressing a non-targeting shRNA (Ctr) or shRNA sequences targeting FGFR3 (sh2, sh3, sh4). Five days after infection cells were plated at density of $2X10^4$ cells/well in triplicate and the number of trypan blue excluding cells was scored at the indicated times. Infection with lentivirus expressing sh-3 and sh-4, the most efficient FGFR3 silencing sequences reverted the growth rate of FGFR3-TACC3 expressing cultures to levels comparable to those of Rat1A transduced with empty vector. Values are the means ± standard deviation (n = 3). At th right sided figure, GSC-1123 cells were transduced with lentivirus expressing a non-targeting shRNA (sh-Ctr) or lentivirus expressing sh-3 and sh-4 sequences targeting FGFR3. Western Blot analysis was performed on parallel cultures using the FGFR3 antibody to the detect FGFT3-TACC3 fusion protein. β-actin is shown as a control for loading.

[0074] **FIG. 7D** shows that the FGFR inhibitor PD173074 suppresses tumor growth of glioma sub-cutaneous xenografts generated by *Ink4A;Arf-/-* astrocytes expressing FGFR3-TACC3. After tumor establishment (200-300 mm^3, arrow) mice were treated with vehicle or PD173074 (50 mg/kg) for 14 days. Values are mean tumor volumes ± standard error (n = 7 mice per group).

[0075] **FIG. 7E** is a Kaplan-Meier analysis of glioma-bearing mice following orthotopic implantation of *Ink4A;Arf-/-* astrocytes transduced with FGFR3-TACC3. After tumor engraftment mice were treated with vehicle (n=9) or AZD4547 (50 mg/kg) (n=7) for 20 days ($p = 0.001$).

[0076] **FIG. 8** shows a schematic of the TX-Fuse pipeline for the identification of fusion transcripts from RNA-Seq data generated from nine GSC cultures. The continued figure shows a schematic of the Exome-Fuse pipeline for the identification of gene fusion rearrangements from DNA exome sequences of 84 GBM TCGA tumor samples.

[0077] **FIG. 9** shows the validation of fusion transcripts identified by RNA-seq of nine

GSCs. Sanger sequencing chromatograms show the reading frames at the breakpoint and putative translation of the fusion proteins in the positive samples (right side). The left side shows gels of RT-PCR conducted. (**A**) POLR2A-WRAP53. (**B**) CAPZB-UBR4. (**C**) ST8SIA4-PAM. (**D**) PIGU-NCOA6.

[0078] **FIG. 9E** shows the fusion transcripts identified by whole transcriptome sequencing of nine GSCs. 54 split-reads are shown aligning on the breakpoint of the POLR2A-WRAP53 fusion. The predicted reading frame at the breakpoint is shown at the top with POLR2A sequences in red (left) and WRAP53 in blue (right). On the continued page, 48 split-reads are shown aligning on the breakpoint of the CAPZB-UBR4 fusion. The predicted reading frame at the breakpoint is shown at the top with CAPZB sequences in red (left) and UBR4 in blue (right). On the continued page after, 29 split-reads are shown aligning on the breakpoint of the ST8SIA4-PAM fusion. The predicted reading frame at the breakpoint is shown at the top with ST8SIA4 sequences in red (left) and PAM in blue (right). On the subsequent continued page, 17 split-reads are shown (top) aligning on the breakpoint of the PIGU-NCOA6 fusion. The predicted reading frame at the breakpoint is shown at the top with PIGU sequences in red (left) and NCOA6 in blue (right). Also (below), 6 split-reads are shown aligning on the breakpoint of the IFNAR2-IL10RB fusion. The predicted reading frame at the breakpoint is shown at the top with IFNAR2 sequences in red (left) and IL10RB in blue (right).

[0079] **FIG. 10A** shows the analysis and validation of the expression of fused transcripts in GSCs and GBM sample. Expression measured by read depth from RNA-seq data. Light grey arcs indicate predicted components of transcripts fused together. Overall read depth (blue; "grey" in black and white image) and split insert depth (red; "dark grey" in black and white image) are depicted in the graph, with a 50-read increment and a maximum range of 1800 reads. Note the very high level of expression in the regions of the genes implicated in the fusion events, particularly for FGFR3-TACC3.

[0080] **FIG. 10B** shows the analysis and validation of the expression of fused transcripts in GSCs and GBM sample. Top panel, qRT-PCR showing the very high expression of FGFR3 and TACC3 mRNA sequences included in the FGFR3-TACC3 fusion transcript in GSC-1123. Bottom panel, for comparison the expression of sequences of WRAP53 mRNA included in the POLR2A-WRAP53 fusion in GSC-0114 is also shown.

[0081] **FIG. 10C** shows the expression of the FGFR3-TACC3 protein in GSC-1123 and

GBM-1123. Western blot analysis with a monoclonal antibody, which recognizes the N-terminal region of human FGFR3 shows expression of a ~150 kD protein in GSC-1123 but not in the GSC cultures GSC-0331 and GSC-0114, which lack the FGFR3-TACC3 rearrangement.

[0082] **FIG. 10D** shows the analysis and validation of the expression of fused transcripts in GSCs and GBM sample. Immunostaining analysis with the FGFR3 antibody of the tumor GBM-1123 (top panel) and a GBM tumor lacking the FGFR3-TACC3 rearrangement. FGFR3 (red; "light grey" in black and white image), DNA (DAPI, blue; "grey" in black and white image). The pictures were taken at low (left) and high (right) magnification.

[0083] **FIG. 10E** shows MS/MS analysis of the ~150 kD fusion protein immunoprecipitated by the monoclonal anti-FGFR3 antibody from GSC-1123, identifying three unique peptides mapping to the FGFR3 (FGFR3 Peptide 1, 2, and 3) and three peptides mapping to the C-terminal region of TACC3 (TACC Peptide 1, 2, and 3).

[0084] **FIGS. 11A-C** shows Rat1A cells transduced with control lentivirus or lentivurus expressing FGFR3, TACC3, FGFR3-TACC3 (**FIG. 11A**) that were analyzed by Western blot with an antibody recognizing the N-terminus of FGFR3 (included in the FGFR3-TACC3 fusion protein) or the N-terminus of TACC3 (not included in the FGFR3-TACC3 fusion protein). **FIG. 11B** shows quantitative Western blot analysis of endogenous FGFR3-TACC3 in GSC-1123 compared with lentivirally expressed FGFR3-TACC3 in Rat1A. **FIG. 11C** shows Western blot analysis of FGFR3-TACC3 and FGFR3-TACC3-K508M in Rat1A. α-tubulin is shown as a control for loading.

[0085] **FIGS. 11D-F** shows expression analyses of FGFR3-TACC3 fusion construct (**FIG. 11D**) FGFR3 immunostaining of GBM-1123 (left, upper panel), BTSC1123 (right, upper panel), mouse GBM induced by FGFR3-TACC3 expressing lentivirus (left, lower panel), and sub-cutaneous xenograft of mouse astrocytes transformed by FGFR3-TACC3 fusion (right, lower panel); FGFR3-TACC3, red ("light grey" in black and white image); DNA (DAPI), blue ("grey" in black and white image). **FIG. 11E** shows quantification of FGFR3-TACC3 positive cells in the tumors and cultures of cells shown in **FIG. 11D**. **FIG. 11F** shows a quantitative Western blot analysis of ectopic FGFR3-TACC3 fusion protein in mouse astrocytes and FGFR3-TACC3 induced mouse GBM (mGBM-15 and mGBM-17) compared with the endogenous expression in GBM1123. β-actin is shown as a control for loading. F3-T3: FGFR3-TACC3. α-tubulin or β-actin is shown as a control for loading.

[0086] **FIG. 12A** shows a western blot. *Ink4A;Arf-/-* astrocytes transduced with empty lentivirus or a lentivirus expressing FGFR3-TACC3 were starved of mitogens and left untreated (time 0) or treated with FGF-2 at concentration of 50 ng/ml for the indicated times. Phospho-proteins and total proteins were analyzed by Western blot using the indicated antibodies. α-tubulin is shown as a control for loading.

[0087] **FIGS. 12B** show western blots. *Ink4A;Arf-/-* astrocytes transduced with empty lentivirus or a lentivirus expressing FGFR3-TACC3 or FGFR3-TACC3-K508M were starved of mitogens and left untreated (time 0) or treated for 10 min with FGF-1 at the indicated concentrations. Phospho-proteins and total proteins were analyzed by Western blot using the indicated antibodies. β-actin is shown as a control for loading.

[0088] **FIG. 12C** show western blots. *Ink4A;Arf-/-* astrocytes transduced with empty lentivirus or a lentivirus expressing FGFR3-TACC3 or FGFR3-TACC3-K508M were starved of mitogens and left untreated (time 0) or treated for 10 min with FGF-8 at the indicated concentrations. Phospho-proteins and total proteins were analyzed by Western blot using the indicated antibodies. β-actin is shown as a control for loading.

[0089] **FIGS. 12D-F** shows mitotic localization of FGFR3-TACC3 fusion protein. **FIG. 12D** shows maximum intensity projection confocal image of a representative FGFR3-TACC3 expressing *Ink4A;Arf-/-* mouse astrocyte at metaphase immunostained using the FGFR3 antibody (red; "dark grey" in black and white image). FGFR3-TACC3 displays asymmetric localization on top of one spindle pole. **FIG. 12E** shows maximum intensity projection confocal image of a representative TACC3 expressing *Ink4A;Arf-/-* mouse astrocyte at metaphase immunostained with the TACC3 antibody (red; ("dark grey" in black and white image). TACC3 staining coincides with the spindle microtubules. **FIG. 12F** shows maximum intensity projection confocal image of a representative FGFR3 expressing *Ink4A;Arf-/-* mouse astrocyte at metaphase immunostained with the FGFR3 antibody (red; ("dark grey" in black and white image). FGFR3 does not show a specific staining pattern in mitosis. Cells were co-immunostained using α-tubulin (green; ("light grey" in black and white image) to visualize the mitotic spindle. DNA was counterstained with DAPI (blue; ("grey" in black and white image). Images were acquired at 0.250 μm intervals. Endogenous levels of FGFR3 or TACC3 were undetectable under the applied experimental conditions. F3-T3: FGFR3-TACC3.

[0090] **FIG. 13A** shows sthat the FGFR3-TACC3 protein induces chromosomal mis-

segregation, chromatid cohesion defects and defective spindle checkpoint. Quantitative analysis of metaphase spreads for chromosome segregation defects in *Ink4A;ARF-/-* astrocytes expressing vector control or FGFR3-TACC3 (upper panel). Microscope imaging analysis of chromosome segregation defects in *Ink4A;Arf-/-* astrocytes expressing FGFR3-TACC3 or vector control. Representative images of cells with chromosome missegregation. Arrows point to chromosome misalignments, lagging chromosomes and chromosome bridges.

[0091] **FIGS. 13B-D** shows representative images of premature sister chromatid separation (PMSCS) in *Ink4A;Arf-/-* astrocytes (**FIG. 13B**) and Rat1A cells (**FIG. 13C**) expressing FGFR3-TACC3. Left, panels show representative metaphase spreads. Right, quantitative analysis of metaphases with loss of sister chromatid cohesion. The number of mitosis with PMSCS in *Ink4A;Arf-/-* astrocytes was scored in at least 100 methaphases for each condition in three independent experiments. The number of mitosis with PMSCS was scored in triplicate samples of Rat1A cells. **FIG. 13D** is a graph showing nocodazole was added for the indicated durations to Rat1A-H2B-GFP cells transduced with the specified lentiviruses. The mitotic index at each time point was determined by quantitating the H2B-GFP-positive cells in mitosis at each time point. Data are presented as average and standard deviation (n = 3). F3-T3: FGFR3-TACC3.

[0092] **FIGS. 14A-B** shows growth curves of human primary astrocytes transduced with lentivirus expressing FGFR3-TACC3 fusion or the empty vector. An analysis was conducted of FGFR3-TACC3 fusion mediated growth alteration and specific effect of RTK inhibitors on cells carrying FGFR-TACC fusions. **FIG. 14A** is a graph that shows cell proliferation of human primary astrocytes transduced with lentivirus expressing FGFR3-TACC3 fusion or the empty vector was determined by the MTT assay 7 days after infection (passage 1). Values are the means ± standard deviation (n = 4). *p*-value: 0.0033. **FIG. 14B** is a graph that shows cell proliferation of human primary astrocytes transduced with lentivirus expressing FGFR3-TACC3 fusion or the empty vector was determined by the MTT assay six weeks after the infection (passage 10). Values are the means±standard deviation (n=4). *p*-value: 0.0018.

[0093] **FIGS. 14C-D** shows specific growth inhibitory effect by FGFR inhibitors on FGFR-TACC fusion expressing cells. Cell growth was determined by MTT assay. Rat1A cells transduced with the indicated lentivirus were treated for three days with BGJ398 **(FIG. 14C)** or AZD4547 **(FIG. 14D)** at the indicated concentration. Values are the means ± standard error (n = 4).

[0094] **FIG. 14E** shows the growth inhibitory effect of silencing FGFR3-TACC3 fusion. **(left)** GSC-1123 cells were transduced in triplicate with lentivirus expressing a non-targeting shRNA (Ctr) or lentivirus expressing sh-3 and sh-4 sequences targeting FGFR3. Five days after infection cells were plated at density of $2X10^4$ cells/well in triplicate and the number of trypan blue excluding cells was scored at the indicated times. Values are the means ± standard deviation (n = 3). **(right)** Western Blot analysis was performed on parallel cultures collected five days after infection using the FGFR3 antibody to the detect FGFT3-TACC3 fusion protein. β-actin is shown as a control for loading. (**: *p*-value = < 0.005; ***: *p*-value = < 0.0001).

[0095] **FIG. 15** shows a survival plot of cells treated with PD173074 , NVP-BGJ398, or AZD4547.

[0096] **FIG. 16** shows an *FGFR3-TACC3* gene fusion identified by whole transcriptome sequencing of GSCs. The histogram describes the absolute frequency of each forward and reverse sequence read spanning the breakpoint.

[0097] **FIG. 17** shows transforming activity of FGFR3-TACC3. FGFR3-TACC3 induces anchorage-independent growth in Rat1A fibroblasts (top panels) and a transformed phenotype in *Ink4A;Arf-/-* primary astrocytes (bottom panels).

[0098] **FIG. 18** shows transforming activity of FGFR3-TACC3. Kaplan-Meier survival curves of mice injected intracranially with pTomo-shp53 (n = 8), pTomo-FGFR3-TACC3-shp53 (n = 8) and pTomo-EGFRvIII-shp53 (n = 7) are shown. Points on the curves indicate deaths (log-rank test, p = 0.025, pTomo-shp53 vs. pTomo-FGFR3-TACC3-shp53).

[0099] **FIG. 19** shows that inhibition of FGFR-TK activity corrects the aneuploidy and suppresses tumor growth initiated by FGFR3-TACC3. Short-term growth inhibition assays are shown of Rat1A transduced with the indicated lentivirus and treated with PD173470 at the indicated concentrations. Cells were treated for three days. Cell viability was determined by the MTT assay. Error bars show means ± standard error (n = 4).

[00100] **FIG. 20** is a growth inhibition assay of human astrocytes transduced with the indicated lentivirus and treated for four days with PD173470 at the indicated concentration. Cell viability was determined by the MTT assay. Error bars show means ± standard error (n = 4).

[00101] **FIG. 21** is a graph showing a growth inhibition assay of human astrocytes

transduced with the indicated lentivirus and treated for four days with PD173470 at the indicated concentration. Cell viability was determined by the MTT assay. Error bars show means ± standard error (n = 4).

[00102] **FIG. 22** shows graphs of the survival of Rat1A cells in short-term growth inhibition assays. (**Top graph**) Rat1A cells were transduced with the indicated ptomo constructs and treated with PD173074 at the indicated concentrations . Cells were treated for three days. Cell viability was determined by the MTT assay. Error bars show means ± standard error (n = 4). In the bottom panel, a western blot photograph is shown.

[00103] **FIG. 23** shows that inhibition of FGFR-TK activity corrects the aneuploidy and suppresses tumor growth initiated by FGFR3-TACC3. A plot is shown of karyotype analysis of Rat1A cells transduced with control or FGFR3-TACC3 lentivirus and treated with vehicle (DMSO) or PD173470 (100 nM) for five days.

[00104] **FIG. 24** shows Survival of glioma-bearing mice was tracked following intracranial implantation of *Ink4A;Arf-/-* astrocytes transduced with FGFR3-TACC3. After tumor engraftment mice were treated with vehicle or AZD4547 (50 mg/kg) for 20 days (vehicle, n = 7; AZD4547, n = 6; p = 0.001).

[00105] **FIG. 25** shows the position of the peptides from **FIG. 10E** in the amino acid sequence of the FGFR3-TACC3 fusion protein, which are highlighted in pink (FGFR3; underlined) and blue (TACC3; dotted lines).

[00106] **FIG. 26** shows Kaplan-Meier analysis of IDH mutant and *FGFR3-TACC3* positive human GBM. Log rank test *p*-value: 0.0169.

[00107] **FIG. 27** is a picture that shows tumor xenografts that were induced following sub-cutaneous injection of *Ink4A;Arf-/-* mouse astrocytes transduced with lentivirus expressing FGFR3-TACC3 (upper panel A, right flank) or FGFR1-TACC1 (lower panel B, right flank) fusion, but not with the empty vector (upper panel, left flank) or FGFR3-TACC3 carrying a K508M mutation in the kinase domain (FGFR3-TACC3-K508M; lower panel, left flank).

[00108] **FIG. 28** shows constitutive auto-phosphorylation of FGFR3-TACC3 fusion. BTSC derived from FGFR3-TACC3 or RasV12 induced mouse GBM were left untreated or treated with 500nM PD173074 for the indicated times. Phospho-proteins and total proteins were analyzed by Western blot using the indicated antibodies. β-actin is shown as a control for loading.

[00109] **FIG. 29** shows Z-stacked confocal images of the representative FGFR3-TACC3 expressing *Ink4A;Arf-/-* mouse astrocyte shown as a maximum intensity projection. Cells were immunostained using FGFR3 (red; "dark grey" in black and white image) and α-tubulin (green; ("light grey" in black and white image). DNA was counterstained with DAPI (blue; ("grey" in black and white image). Images were acquired at 0.250 μm intervals. Coordinates of the image series are indicated. F3-T3: FGFR3-TACC3.

[00110] **FIG. 30** shows examples of SKY karyotype analysis painting two different cells from the same culture of GSC-1123, illustrating the ongoing CIN and aneuploidy. Details of the karyotype analysis of 20 cells are reported in Table 6.

[00111] **FIG. 31** is a graphical representation of segmented CNVs data visualized using the Integrated Genomic Viewers software. Three bladder Urothelial Carcinoma harbor FGFR3-TACC3 gene fusions (black box). Red indicates amplification (A), blue indicates deletion (D).

[00112] **FIG. 32** is a graphical representation of segmented CNVs data visualized using the Integrated Genomic Viewers software. One Breast Carcinoma harbors FGFR3-TACC3 gene fusions (black box). Red indicates amplification (A), blue indicates deletion (D).

[00113] **FIG. 33** is a graphical representation of segmented CNVs data visualized using the Integrated Genomic Viewers software. One Colorectal Carcinoma harbors FGFR3-TACC3 gene fusions (black box). Red indicates amplification (A), blue indicates deletion (D).

[00114] **FIG. 34** is a graphical representation of segmented CNVs data visualized using the Integrated Genomic Viewers software. One Lung Squamous Cell Carcinoma harbors FGFR3-TACC3 gene fusions (black box). Red indicates amplification (A), blue indicates deletion (D).

[00115] **FIG. 35** is a graphical representation of segmented CNVs data visualized using the Integrated Genomic Viewers software. One Head abd Neck Squamous Cell Carcinoma harbors FGFR3-TACC3 gene fusions (black box). Red indicates amplification (A), blue indicates deletion (D).

DETAILED DESCRIPTION OF THE INVENTION

[00116] Glioblastoma multiformes (GBMs) are the most common form of brain tumors in adults accounting for 12-15% of intracranial tumors and 50-60% of primary brain tumors.

GBM is among the most lethal forms of human cancer. The history of successful targeted therapy of cancer largely coincides with the inactivation of recurrent and oncogenic gene fusions in hematological malignancies and recently in some types of epithelial cancer. GBM is among the most lethal and incurable forms of human cancer. Targeted therapies against common genetic alterations in GBM have not changed the dismal clinical outcome of the disease, most likely because they have systematically failed to eradicate the truly addicting oncoprotein activities of GBM. Recurrent chromosomal rearrangements resulting in the creation of oncogenic gene fusions have not been found in GBM.

[00117] GBM is among the most difficult forms of cancer to treat in humans (*1*). So far, the therapeutic approaches that have been tested against potentially important oncogenic targets in GBM have met limited success (*2-4*). Recurrent chromosomal translocations leading to production of oncogenic fusion proteins are viewed as initiating and addicting events in the pathogenesis of human cancer, thus providing the most desirable molecular targets for cancer therapy (*5, 6*). Recurrent and oncogenic gene fusions have not been found in GBM. Chromosomal rearrangements are hallmarks of hematological malignancies but recently they have also been uncovered in subsets of solid tumors (breast, prostate, lung and colorectal carcinoma) (*7, 8*). Important and successful targeted therapeutic interventions for patients whose tumors carry these rearrangements have stemmed from the discovery of functional gene fusions, especially when the translocations involve kinase-coding genes (*BCR-ABL, EML4-ALK*) (*9, 10*).

[00118] A hallmark of GBM is rampant chromosomal instability (CIN), which leads to aneuploidy (*11*). CIN and aneuploidy are early events in the pathogenesis of cancer (*12*). It has been suggested that genetic alterations targeting mitotic fidelity might be responsible for missegregation of chromosomes during mitosis, resulting in aneuploidy (*13, 14*).

[00119] Fibroblast growth factor receptors (FGFR) are transmembrane receptors that bind to members of the fibroblast growth factor family of proteins. The structure of the FGFRs consist of an extracellular ligand binding domain comprised of three Ig-like domains, a single transmembrane helix domain, and an intracellular domain with tyrosine kinase activity (Johnson, D.E., Williams, E. T. Structural and functional diversity in the FGF receptor multigene family. (1993) *Adv. Cancer Res*, 60:1–41).

[00120] Transforming acidic coiled-coiled protein (TACC) stabilize microtubules during

mitosis by recruiting minispindles (Msps)/XMAP215 proteins to centrosomes. TACCs have been implicated in cancer.

[00121] From a medical perspective, the FGFR-TACC fusions provide the first "bona-fide" oncogenically addictive gene fusions in GBM whose identification has long been overdue in this disease.

DNA and AminoAcid Manipulation Methods and Purification Thereof

[00122] The practice of aspects of the present invention can employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. *See*, e.g., Molecular Cloning A Laboratory Manual, 3rd Ed., ed. by Sambrook (2001), Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1989); DNA Cloning, Volumes I and II (D. N. Glover ed., 1985); Oligonucleotide Synthesis (M. J. Gait ed., 1984); Mullis et al. U.S. Pat. No: 4,683,195; Nucleic Acid Hybridization (B. D. Hames & S. J. Higgins eds. 1984); Transcription and Translation (B. D. Hames & S. J. Higgins eds. 1984); Culture Of Animal Cells (R. I. Freshney, Alan R. Liss, Inc., 1987); Immobilized Cells and Enzymes (IRL Press, 1986); B. Perbal, A Practical Guide To Molecular Cloning (1984); the series, Methods In Enzymology (Academic Press, Inc., N.Y.), specifically, Methods In Enzymology, Vols. 154 and 155 (Wu et al. eds.); Gene Transfer Vectors For Mammalian Cells (J. H. Miller and M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); Immunochemical Methods In Cell And Molecular Biology (Caner and Walker, eds., Academic Press, London, 1987); Handbook Of Experimental Immunology, Volumes I-IV (D. M. Weir and C. C. Blackwell, eds., 1986); Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). All patents, patent applications and references cited herein are incorporated by reference in their entireties.

[00123] One skilled in the art can obtain a protein in several ways, which include, but are not limited to, isolating the protein via biochemical means or expressing a nucleotide sequence encoding the protein of interest by genetic engineering methods.

[00124] A protein is encoded by a nucleic acid (including, for example, genomic DNA, complementary DNA (cDNA), synthetic DNA, as well as any form of corresponding RNA). For example, it can be encoded by a recombinant nucleic acid of a gene. The proteins of the

invention can be obtained from various sources and can be produced according to various techniques known in the art. For example, a nucleic acid that encodes a protein can be obtained by screening DNA libraries, or by amplification from a natural source. A protein can be a fragment or portion thereof. The nucleic acids encoding a protein can be produced via recombinant DNA technology and such recombinant nucleic acids can be prepared by conventional techniques, including chemical synthesis, genetic engineering, enzymatic techniques, or a combination thereof. For example, a fusion protein of the invention comprises a tyrosine kinase domain of an FGFR protein fused to a polypeptide that constitutively activates the tyrosine kinase domain of the FGFR protein. For example, a fusion protein of the invention comprises a transforming acidic coiled-coil (TACC) domain fused to a polypeptide with a tyrosine kinase domain, wherein the TACC domain constitutively activates the tyrosine kinase domain. An example of a FGFR1-TACC1 polypeptide has the amino acid sequence shown in SEQ ID NO: 150. An example of a FGFR3-TACC3 protein is the polypeptide encoded by the nucleic acid having the nucleotide sequence shown in SEQ ID NO: 94. Examples of a FGFR3-TACC3 polypeptide has the amino acid sequence shown in SEQ ID NO: 79, 158, 159, 160, or 161.

[00125] The Genbank ID for the FGFR3 gene is 2261. Three isoforms are listed for FGFGR3, e.g., having Genebank Accession Nos. NP_000133 (corresponding nucleotide sequence NM_000142); NP_001156685 (corresponding nucleotide sequence NM_001163213); NP_075254 (corresponding nucleotide sequence NM_022965).

[00126] SEQ ID NO: 90 is the FGFR3 Amino Acid Sequence, Transcript Variant 1 (NP_000133; 806 aa).

```
  1 MGAPACALAL CVAVAIVAGA SSESLGTEQR VVGRAAEVPG PEPGQQEQLV FGSGDAVELS
 61 CPPPGGGPMG PTVWVKDGTG LVPSERVLVG PQRLQVLNAS HEDSGAYSCR QRLTQRVLCH
121 FSVRVTDAPS SGDDEDGEDE AEDTGVDTGA PYWTRPERMD KKLLAVPAAN TVRFRCPAAG
181 NPTPSISWLK NGREFRGEHR IGGIKLRHQQ WSLVMESVVP SDRGNYTCVV ENKFGSIRQT
241 YTLDVLERSP HRPILQAGLP ANQTAVLGSD VEFHCKVYSD AQPHIQWLKH VEVNGSKVGP
301 DGTPYVTVLK TAGANTTDKE LEVLSLHNVT FEDAGEYTCL AGNSIGFSHH SAWLVVLPAE
361 EELVEADEAG SVYAGILSYG VGFFLFILVV AAVTLCRLRS PPKKGLGSPT VHKISRFPLK
421 RQVSLESNAS MSSNTPLVRI ARLSSGEGPT LANVSELELP ADPKWELSRA RLTLGKPLGE
481 GCFGQVVMAE AIGIDKDRAA KPVTVAVKML KDDATDKDLS DLVSEMEMMK MIGKHKNIIN
541 LLGACTQGGP LYVLVEYAAK GNLREFLRAR RPPGLDYSFD TCKPPEEQLT FKDLVSCAYQ
601 VARGMEYLAS QKCIHRDLAA RNVLVTEDNV MKIADFGLAR DVHNLDYYKK TTNGRLPVKW
661 MAPEALFDRV YTHQSDVWSF GVLLWEIFTL GGSPYPGIPV EELFKLLKEG HRMDKPANCT
721 HDLYMIMREC WHAAPSQRPT FKQLVEDLDR VLTVTSTDEY LDLSAPFEQY SPGGQDTPSS
781 SSSGDDSVFA HDLLPPAPPS SGGSRT
```

[00127] SEQ ID NO: 91 is the FGFR3 Nucleotide Sequence, Transcript Variant 1

(NM_000142; 4304 bp).

```
   1 gtcgcgggca gctggcgccg cgcggtcctg ctctgccggt cgcacggacg caccggcggg
  61 ccgccggccg gagggacggg gcgggagctg ggcccgcgga cagcgagccg gagcgggagc
 121 cgcgcgtagc gagccgggct ccggcgctcg ccagtctccc gagcggcgcc cgcctcccgc
 181 cggtgcccgc gccgggccgt gggggggcagc atgcccgcgc gcgctgcctg aggacgccgc
 241 ggcccccgcc cccgccatgg gcgcccctgc ctgcgccctc gcgctctgcg tggccgtggc
 301 catcgtggcc ggcgcctcct cggagtcctt ggggacggag cagcgcgtcg tggggcgagc
 361 ggcagaagtc ccgggcccag agcccggcca gcaggagcag ttggtcttcg gcagcgggga
 421 tgctgtggag ctgagctgtc ccccgcccgg gggtggtccc atggggccca ctgtctgggt
 481 caaggatggc acagggctgg tgccctcgga gcgtgtcctg gtggggcccc agcggctgca
 541 ggtgctgaat gcctcccacg aggactccgg ggcctacagc tgccggcagc ggctcacgca
 601 gcgcgtactg tgccacttca gtgtgcgggt gacagacgct ccatcctcgg gagatgacga
 661 agacggggag gacgaggctg aggacacagg tgtggacaca ggggcccctt actggacacg
 721 gcccgagcgg atggacaaga agctgctggc cgtgccggcc gccaacaccg tccgcttccg
 781 ctgcccagcc gctggcaacc ccactccctc catctcctgg ctgaagaacg gcagggagtt
 841 ccgcggcgag caccgcattg gaggcatcaa gctgcggcat cagcagtgga gcctggtcat
 901 ggaaagcgtg gtgccctcgg accgcggcaa ctacacctgc gtcgtggaga acaagtttgg
 961 cagcatccgg cagacgtaca cgctggacgt gctggagcgc tccccgcacc ggcccatcct
1021 gcaggcgggg ctgccggcca accagacggc ggtgctgggc agcgacgtgg agttccactg
1081 caaggtgtac agtgacgcac agccccacat ccagtggctc aagcacgtgg aggtgaatgg
1141 cagcaaggtg ggcccggacg gcacacccta cgttaccgtg ctcaagacgg cgggcgctaa
1201 caccaccgac aaggagctag aggttctctc cttgcacaac gtcacctttg aggacgccgg
1261 ggagtacacc tgcctggcgg gcaattctat tgggttttct catcactctg cgtggctggt
1321 ggtgctgcca gccgaggagg agctggtgga ggctgacgag gcgggcagtg tgtatgcagg
1381 catcctcagc tacggggtgg gcttcttcct gttcatcctg gtggtggcgg ctgtgacgct
1441 ctgccgcctg cgcagccccc ccaagaaagg cctgggctcc cccaccgtgc acaagatctc
1501 ccgcttcccg ctcaagcgac aggtgtccct ggagtccaac gcgtccatga gctccaacac
1561 accactggtg cgcatcgcaa ggctgtcctc aggggagggc cccacgctgg ccaatgtctc
1621 cgagctcgag ctgcctgccg accccaaatg ggagctgtct cgggcccggc tgaccctggg
1681 caagcccctt ggggagggct gcttcggcca ggtggtcatg gcggaggcca tcggcattga
1741 caaggaccgg gccgccaagc ctgtcaccgt agccgtgaag atgctgaaag acgatgccac
1801 tgacaaggac ctgtcggacc tggtgtctga gatggagatg atgaagatga tcgggaaaca
1861 caaaaacatc atcaacctgc tgggcgcctg cacgcagggc gggccctgt acgtgctggt
1921 ggagtacgcg gccaagggta acctgcggga gtttctgcgg gcgcggcggc ccccgggcct
1981 ggactactcc ttcgacacct gcaagccgcc cgaggagcag ctcaccttca aggacctggt
2041 gtcctgtgcc taccaggtgg cccggggcat ggagtacttg gcctcccaga agtgcatcca
2101 cagggacctg gctgcccgca atgtgctggt gaccgaggac aacgtgatga agatcgcaga
2161 cttcgggctg gcccgggacg tgcacaacct cgactactac aagaagacaa ccaacggccg
2221 gctgcccgtg aagtggatgg cgcctgaggc cttgtttgac cgagtctaca ctcaccagag
2281 tgacgtctgg tcctttgggg tcctgctctg ggagatcttc acgctggggg gctccccgta
2341 ccccggcatc cctgtggagg agctcttcaa gctgctgaag gagggccacc gcatggacaa
2401 gcccgccaac tgcacacacg acctgtacat gatcatgcgg gagtgctggc atgccgcgcc
2461 ctcccagagg cccaccttca agcagctggt ggaggacctg gaccgtgtcc ttaccgtgac
2521 gtccaccgac gagtacctgg acctgtcggc gcctttcgag cagtactccc cgggtggcca
2581 ggacaccccc agctccagct cctcagggga cgactccgtg tttgcccacg acctgctgcc
2641 ccccggcccca cccagcagtg ggggctcgcg gacgtgaagg gccactggtc cccaacaatg
2701 tgaggggtcc ctagcagccc accctgctgc tggtgcacag ccactccccg gcatgagact
2761 cagtgcagat ggagagacag ctacacagag ctttggtctg tgtgtgtgtg tgtgcgtgtg
2821 tgtgtgtgtg tgtgcacatc cgcgtgtgcc tgtgtgcgtg cgcatcttgc ctccaggtgc
2881 agaggtaccc tgggtgtccc cgctgctgtg caacggtctc ctgactggtg ctgcagcacc
2941 gaggggcctt tgttctgggg ggacccagtg cagaatgtaa gtgggcccac ccggtgggac
3001 ccccgtgggg cagggagctg ggcccgacat ggctccggcc tctgcctttg caccacggga
3061 catcacaggg tgggcctcgg cccctcccac acccaaagct gagcctgcag ggaagcccca
3121 catgtccagc accttgtgcc tggggtgtta gtggcaccgc ctccccacct ccaggctttc
3181 ccacttccca ccctgccccc cagagactga aattacgggt acctgaagat gggagccttt
3241 acctttttatg caaaaggttt attccggaaa ctagtgtaca tttctataaa tagatgctgt
3301 gtatatggta tatatacata tatatatata acatatatgg aagaggaaaa ggctggtaca
3361 acggaggcct gcgaccctgg gggcacagga ggcaggcatg gccctgggcg gggcgtgggg
3421 gggcgtggag ggaggcccca gggggtctca cccatgcaag cagaggacca gggcctttc
```

```
3481 tggcaccgca gttttgtttt aaaactggac ctgtatattt gtaaagctat ttatgggccc
3541 ctggcactct tgttcccaca ccccaacact tccagcattt agctggccac atggcggaga
3601 gttttaattt ttaacttatt gacaaccgag aaggtttatc ccgccgatag agggacggcc
3661 aagaatgtac gtccagcctg ccccggagct ggaggatccc ctccaagcct aaaaggttgt
3721 taatagttgg aggtgattcc agtgaagata ttttatttcc tttgtccttt ttcaggagaa
3781 ttagatttct ataggatttt tctttaggag atttatttct tggacttcaa agcaagctgg
3841 tattttcata caaattcttc taattgctgt gtgtcccagg cagggagacg gtttccaggg
3901 aggggccggc cctgtgtgca ggttccgatg ttattagatg ttacaagttt atatatatct
3961 atatatataa tttattgagt ttttacaaga tgtatttgtt gtagacttaa cacttcttac
4021 gcaatgcttc tagagtttta tagcctggac tgctaccttt caaagcttgg agggaagccg
4081 tgaattcagt tggttcgttc tgtactgtta ctgggccctg agtctgggca gctgtccctt
4141 gcttgcctgc agggccatgg ctcagggtgg tctcttcttg gggcccagtg catggtggcc
4201 agaggtgtca cccaaaccgg caggtgcgat tttgttaacc cagcgacgaa ctttccgaaa
4261 aataaagaca cctggttgct aacctggaaa aaaaaaaaaa aaaa
```

[00128] The Genbank ID for the TACC3 gene is 10460. SEQ ID NO: 92 is the TACC3 Amino Acid Sequence (NP_006333) (838 aa).

```
  1 MSLQVLNDKN VSNEKNTENC DFLFSPPEVT GRSSVLRVSQ KENVPPKNLA KAMKVTFQTP
 61 LRDPQTHRIL SPSMASKLEA PFTQDDTLGL ENSHPVWTQK ENQQLIKEVD AKTTHGILQK
121 PVEADTDLLG DASPAFGSGS SSESGPGALA DLDCSSSSQS PGSSENQMVS PGKVSGSPEQ
181 AVEENLSSYS LDRRVTPASE TLEDPCRTES QHKAETPHGA EEECKAETPH GAEEECRHGG
241 VCAPAAVATS PPGAIPKEAC GGAPLQGLPG EALGCPAGVG TPVPADGTQT LTCAHTSAPE
301 STAPTNHLVA GRAMTLSPQE EVAAGQMASS SRSGPVKLEF DVSDGATSKR APPPRRLGER
361 SGLKPPLRKA AVRQQKAPQE VEEDDGRSGA GEDPPMPASR GSYHLDWDKM DDPNFIPFGG
421 DTKSGCSEAQ PPESPETRLG QPAAEQLHAG PATEEPGPCL SQQLHSASAE DTPVVQLAAE
481 TPTAESKERA LNSASTSLPT SCPGSEPVPT HQQGQPALEL KEESFRDPAE VLGTGAEVDY
541 LEQFGTSSFK ESALRKQSLY LKFDPLLRDS PGRPVPVATE TSSMHGANET PSGRPREAKL
601 VEFDFLGALD IPVPGPPPGV PAPGGPPLST GPIVDLLQYS QKDLDAVVKA TQEENRELRS
661 RCEELHGKNL ELGKIMDRFE EVVYQAMEEV QKQKELSKAE IQKVLKEKDQ LTTDLNSMEK
721 SFSDLFKRFE KQKEVIEGYR KNEESLKKCV EDYLARITQE GQRYQALKAH AEEKLQLANE
781 EIAQVRSKAQ AEALALQASL RKEQMRIQSL EKTVEQKTKE NEELTRICDD LISKMEKI
```

[00129] SEQ ID NO: 93 is the TACC3 Nucleotide Sequence (NM_006342) (2847 bp):

```
   1 gcgtttgaaa ctccggcgcg ccggcggcca tcaagggcta gaagcgcgac ggcggtagca
  61 gctaggcttg gccccccggcg tggagcagac gcggaccccct ccttcctggc ggcggcggcg
 121 cgggctcaga gcccggcaac gggcgggcgg gcagaatgag tctgcaggtc ttaaacgaca
 181 aaaatgtcag caatgaaaaa aatacagaaa attgcgactt cctgttttcg ccaccagaag
 241 ttaccggaag atcgtctgtt cttcgtgtgt cacagaaaga aaatgtgcca cccaagaacc
 301 tggccaaagc tatgaaggtg acttttcaga cacctctgcg ggatccacag acgcacagga
 361 ttctaagtcc tagcatggcc agcaaacttg aggctccttt cactcaggat gacacccttg
 421 gactggaaaa ctcacacccg gtctggacac agaaagagaa ccaacagctc atcaaggaag
 481 tggatgccaa aactactcat ggaattctac agaaaccagt ggaggctgac accgacctcc
 541 tgggggatgc aagcccagcc tttgggagtg gcagctccag cgagtctggc ccaggtgccc
 601 tggctgacct ggactgctca agctcttccc agagcccagg aagttctgag aaccaaatgg
 661 tgtctccagg aaaagtgtct ggcagccctg agcaagccgt ggaggaaaac cttagttcct
 721 attccttaga cagaagagtg acacccgcct ctgagaccct agaagaccct tgcaggacag
 781 agtcccagca caaagcggag actccgcacg gagccgagga agaatgcaaa gcggagactc
 841 cgcacggagc cgaggaggaa tgccggcacg gtggggtctg tgctcccgca gcagtggcca
 901 cttcgcctcc tggtgcaatc cctaaggaag cctgcggagg agcacccctg cagggtctgc
 961 ctggcgaagc cctgggctgc cctgcgggtg tgggcacccc cgtgccagca gatggcactc
1021 agacccttac ctgtgcacac acctctgctc ctgagagcac agccccaacc aaccacctgg
1081 tggctggcag ggccatgacc ctgagtcctc aggaagaagt ggctgcaggc caaatggcca
1141 gctcctcgag gagcggacct gtaaaactag aatttgatgt atctgatggc gccaccagca
1201 aaagggcacc cccaccaagg agactgggag agaggtccgg cctcaagcct cccttgagga
1261 aagcagcagt gaggcagcaa aaggccccgc aggaggtgga ggaggacgac ggtaggagcg
```

1321 gagcaggaga ggaccccccc atgccagctt ctcggggctc ttaccacctc gactgggaca
1381 aaatggatga cccaaacttc atcccgttcg gaggtgacac caagtctggt tgcagtgagg
1441 cccagccccc agaaagccct gagaccaggc tgggccagcc agcggctgaa cagttgcatg
1501 ctgggcctgc cacggaggag ccaggtccct gtctgagcca gcagctgcat tcagcctcag
1561 cggaggacac gcctgtggtg cagttggcag ccgagacccc aacagcagag agcaaggaga
1621 gagccttgaa ctctgccagc acctcgcttc ccacaagctg tccaggcagt gagccagtgc
1681 ccacccatca gcagggggcag cctgccttgg agctgaaaga ggagagcttc agagaccccg
1741 ctgaggttct aggcacgggc gcggaggtgg attacctgga gcagtttgga acttcctcg**t**
1801 **ttaaggagtc ggccttgagg aagcagtcct tatacctcaa gttcgacccc ctcctgaggg**
1861 **acagtcctgg tagaccagtg cccgtggcca** ccgagaccag cagcatgcac ggtgcaaatg
1921 agactccctc aggacgtccg cgggaagcca agcttgtgga gttcgatttc ttgggagcac
1981 tggacattcc tgtgccaggc ccacccccag gtgttcccgc gcctgggggc ccacccctgt
2041 ccaccggacc tatagtggac ctgctccagt acagccagaa ggacctggat gcagtggtaa
2101 aggcgacaca ggaggagaac cgggagctga ggagcaggtg tgaggagctc cacgggaaga
2161 acctggaact ggggaagatc atggacaggt tcgaagaggt tgtgtaccag gccatggagg
2221 aagttcagaa gcagaaggaa ctttccaaag ctgaaatcca gaaagttcta aaagaaaaag
2281 accaacttac cacagatctg aactccatgg agaagtcctt ctccgacctc ttcaagcgtt
2341 ttgagaaaca gaaagaggtg atcgagggct accgcaagaa cgaagagtca ctgaagaagt
2401 gcgtggagga ttacctggca aggatcaccc aggagggcca gaggtaccaa gccctgaagg
2461 cccacgcgga ggagaagctg cagctggcaa acgaggagat cgcccaggtc cggagcaagg
2521 cccaggcgga agcgttggcc ctccaggcca gcctgaggaa ggagcagatg cgcatccagt
2581 cgctggagaa gacagtggag cagaagacta aagagaacga ggagctgacc aggatctgcg
2641 acgacctcat ctccaagatg gagaagatct gacctccacg gagccgctgt ccccgccccc
2701 ctgctcccgt ctgtctgtcc tgtctgattc tcttaggtgt catgttcttt tttctgtctt
2761 gtcttcaact tttttaaaaa ctagattgct ttgaaaacat gactcaataa aagtttcctt
2821 tcaatttaaa cactgaaaaa aaaaaaa

[00130] SEQ ID NO: 94 is the nucleotide sequence of FGFR3-TACC3.

1 gtcgcgggca gctggcgccg cgcggtcctg ctctgccggt cgcacggacg caccggcggg
61 ccgccggccg gagggacggg gcgggagctg ggccccgcgga cagcgagccg gagcgggagc
121 cgcgcgtagc gagccgggct ccggcgctcg ccagtctccc gagcggcgcc cgcctcccgc
181 cggtgcccgc gccgggccgt gggggggcagc atgcccgcgc gcgctgcctg aggacgccgc
241 ggcccccgcc cccgccatgg gcgcccctgc ctgcgccctc gcgctctgcg tggccgtggc
301 catcgtggcc ggcgcctcct cggagtcctt ggggacggag cagcgcgtcg tggggcgagc
361 ggcagaagtc ccgggcccag agccccggcca gcaggagcag ttggtcttcg gcagcgggga
421 tgctgtggag ctgagctgtc ccccgccccgg gggtggtccc atggggccca ctgtctgggt
481 caaggatggc acagggctgg tgccctcgga gcgtgtcctg gtggggcccc agcggctgca
541 ggtgctgaat gcctcccacg aggactccgg ggcctacagc tgccggcagc ggctcacgca
601 gcgcgtactg tgccacttca gtgtgcgggt gacagacgct ccatcctcgg gagatgacga
661 agacggggag gacgaggctg aggacacagg tgtggacaca ggggcccctt actggacacg
721 gcccgagcgg atggacaaga agctgctggc cgtgccggcc gccaacaccg tccgcttccg
781 ctgcccagcc gctggcaacc ccactccctc catctcctgg ctgaagaacg gcagggagtt
841 ccgcggcgag caccgcattg gaggcatcaa gctgcggcat cagcagtgga gcctggtcat
901 ggaaagcgtg gtgccctcgg accgcggcaa ctacacctgc gtcgtggaga acaagtttgg
961 cagcatccgg cagacgtaca cgctggacgt gctggagcgc tccccgcacc ggcccatcct
1021 gcaggcgggg ctgccggcca accagacggc ggtgctgggc agcgacgtgg agttccactg
1081 caaggtgtac agtgacgcac agccccacat ccagtggctc aagcacgtgg aggtgaatgg
1141 cagcaaggtg ggcccggacg gcacacccta cgttaccgtg ctcaagacgg cgggcgctaa
1201 caccaccgac aaggagctag aggttctctc cttgcacaac gtcacctttg aggacgccgg
1261 ggagtacacc tgcctggcgg gcaattctat tgggttttct catcactctg cgtggctggt
1321 ggtgctgcca gccgaggagg agctggtgga ggctgacgag gcgggcagtg tgtatgcagg
1381 catcctcagc tacggggtgg gcttcttcct gttcatcctg gtggtggcgg ctgtgacgct
1441 ctgccgcctg cgcagccccc ccaagaaagg cctgggctcc cccaccgtgc acaagatctc
1501 ccgcttcccg ctcaagcgac aggtgtccct ggagtccaac gcgtccatga gctccaacac
1561 accactggtg cgcatcgcaa ggctgtcctc aggggagggc cccacgctgg ccaatgtctc
1621 cgagctcgag ctgcctgccg accccaaatg ggagctgtct cgggcccggc tgaccctggg
1681 caagcccctt ggggagggct gcttcggcca ggtggtcatg gcggaggcca tcggcattga
1741 caaggaccgg gccgccaagc ctgtcaccgt agccgtgaag atgctgaaag acgatgccac
1801 tgacaaggac ctgtcggacc tggtgtctga gatggagatg atgaagatga tcgggaaaca

1861 caaaaacatc atcaacctgc tgggcgcctg cacgcagggc gggcccctgt acgtgctggt
1921 ggagtacgcg gccaagggta acctgcggga gtttctgcgg gcgcgggcgc ccccgggcct
1981 ggactactcc ttcgacacct gcaagccgcc cgaggagcag ctcaccttca aggacctggt
2041 gtcctgtgcc taccaggtgg cccggggcat ggagtacttg gcctcccaga agtgcatcca
2101 cagggacctg gctgcccgca atgtgctggt gaccgaggac aacgtgatga agatcgcaga
2161 cttcgggctg gcccgggacg tgcacaacct cgactactac aagaagacaa ccaacggccg
2221 gctgcccgtg aagtggatgg cgcctgaggc cttgtttgac cgagtctaca ctcaccagag
2281 tgacgtctgg tcctttgggg tcctgctctg ggagatcttc acgctggggg gctccccgta
2341 ccccggcatc cctgtggagg agctcttcaa gctgctgaag gagggccacc gcatggacaa
2401 gcccgccaac tgcacacacg acctgtacat gatcatgcgg ga**gtgctggc atgccgcgcc**
2461 **ctcccagagg cccaccttca agcagctggt ggaggacctg gaccgtgtcc ttaccgtgac**
2521 **gtccaccgac tttaaggagt cggccttgag gaagcagtcc ttatacctca agttcgaccc**
2581 **cctcctgagg gacagtcctg gtagaccagt gcccgtggcc** accgagacca gcagcatgca
2641 cggtgcaaat gagactccct caggacgtcc gcgggaagcc aagcttgtgg agttcgattt
2701 cttgggagca ctggacattc ctgtgccagg cccacccccа ggtgttcccg cgcctggggg
2761 cccacccctg tccaccggac ctatagtgga cctgctccag tacagccaga aggacctgga
2821 tgcagtggta aaggcgacac aggaggagaa ccgggagctg aggagcaggt gtgaggagct
2881 ccacgggaag aacctggaac tggggaagat catggacagg ttcgaagagg ttgtgtacca
2941 ggccatggag gaagttcaga agcagaagga actttccaaa gctgaaatcc agaaaagttct
3001 aaaagaaaaa gaccaactta ccacagatct gaactccatg gagaagtcct tctccgacct
3061 cttcaagcgt tttgagaaac agaaagaggt gatcgagggc taccgcaaga acgaagagtc
3121 actgaagaag tgcgtggagg attacctggc aaggatcacc caggagggcc agaggtacca
3181 agccctgaag gcccacgcgg aggagaagct gcagctggca aacgaggaga tcgcccaggt
3241 ccggagcaag gcccaggcgg aagcgttggc cctccaggcc agcctgagga aggagcagat
3301 gcgcatccag tcgctggaga agacagtgga gcagaagact aaagagaacg aggagctgac
3361 caggatctgc gacgacctca tctccaagat ggagaagatc tgacctccac ggagccgctg
3421 tccccgcccc cctgctcccg tctgtctgtc ctgtctgatt ctcttaggtg tcatgttctt
3481 ttttctgtct tgtcttcaac ttttttaaaa actagattgc tttgaaaaca tgactcaata
3541 aaagtttcct ttcaatttaa acactgaaaa aaaaaaaa

[00131] The Genbank ID for the FGFR1 gene is 2260. Eight isoforms are listed for FGFR1, e.g., having Genebank Accession Nos. NP_001167534 (corresponding nucleotide sequence NM_001174063); NP_001167535 (corresponding nucleotide sequence NM_001174064); NP_001167536 (corresponding nucleotide sequence NM_001174065); NP_001167537 (corresponding nucleotide sequence NM_001174066); NP_001167538 (corresponding nucleotide sequence NM_001174067); NP_056934 (corresponding nucleotide sequence NM_015850); NP_075593 (corresponding nucleotide sequence NM_023105); NP_075594 (corresponding nucleotide sequence NM_023106); NP_075598 (corresponding nucleotide sequence NM_023110).

[00132] SEQ ID NO: 146 is the FGFR1 Amino Acid Sequence for isoform 10, having Genebank Accession No. NP_001167534 (820 aa):

1 MWSWKCLLFW AVLVTATLCT ARPSPTLPEQ AQPWGAPVEV ESFLVHPGDL LQLRCRLRDD
61 VQSINWLRDG VQLAESNRTR ITGEEVEVQD SVPADSGLYA CVTSSPSGSD TTYFSVNVSD
121 ALPSSEDDDD DDDSSSEEKE TDNTKPNRMP VAPYWTSPEK MEKKLHAVPA AKTVKFKCPS
181 SGTPNPTLRW LKNGKEFKPD HRIGGYKVRY ATWSIIMDSV VPSDKGNYTC IVENEYGSIN
241 HTYQLDVVER SPHRPILQAG LPANKTVALG SNVEFMCKVY SDPQPHIQWL KHIEVNGSKI
301 GPDNLPYVQI LKTAGVNTTD KEMEVLHLRN VSFEDAGEYT CLAGNSIGLS HHSAWLTVLE
361 ALEERPAVMT SPLYLEIIIY CTGAFLISCM VGSVIVYKMK SGTKKSDFHS QMAVHKLAKS
421 IPLRRQVSAD SSASMNSGVL LVRPSRLSSS GTPMLAGVSE YELPEDPRWE LPRDRLVLGK

```
481 PLGEGCFGQV VLAEAIGLDK DKPNRVTKVA VKMLKSDATE KDLSDLISEM EMMKMIGKHK
541 NIINLLGACT QDGPLYVIVE YASKGNLREY LQARRPPGLE YCYNPSHNPE EQLSSKDLVS
601 CAYQVARGME YLASKKCIHR DLAARNVLVT EDNVMKIADF GLARDIHHID YYKKTTNGRL
661 PVKWMAPEAL FDRIYTHQSD VWSFGVLLWE IFTLGGSPYP GVPVEELFKL LKEGHRMDKP
721 SNCTNELYMM MRDCWHAVPS QRPTFKQLVE DLDRIVALTS NQEYLDLSMP LDQYSPSFPD
781 TRSSTCSSGE DSVFSHEPLP EEPCLPRHPA QLANGGLKRR
```

[00133] SEQ ID NO: 147 is the FGFR1 Nucleotide Sequence for isoform 10, having Genebank Accession No. NM_001174063 (5895 bp):

```
   1 agatgcaggg gcgcaaacgc caaaggagac caggctgtag gaagagaagg gcagagcgcc
  61 ggacagctcg gcccgctccc cgtcctttgg ggccgcggct ggggaactac aaggcccagc
 121 aggcagctgc agggggcgga ggcggaggag ggaccagcgc gggtgggagt gagagagcga
 181 gccctcgcgc cccgccggcg catagcgctc ggagcgctct tgcggccaca ggcgcggcgt
 241 cctcggcggc gggcggcagc tagcgggagc cgggacgccg gtgcagccgc agcgcgcgga
 301 ggaacccggg tgtgccggga gctgggcggc cacgtccgga cgggaccgag acccctcgta
 361 gcgcattgcg gcgacctcgc cttccccggc cgcgagcgcg ccgctgcttg aaaagccgcg
 421 gaacccaagg acttttctcc ggtccgagct cggggcgccc cgcagggcgc acggtacccg
 481 tgctgcagtc gggcacgccg cggcgccggg gcctccgcag ggcgatggag cccggtctgc
 541 aaggaaagtg aggcgccgcc gctgcgttct ggaggagggg ggcacaaggt ctggagaccc
 601 cgggtggcgg acgggagccc tcccccccgcc ccgcctccgg ggcaccagct ccggctccat
 661 tgttcccgcc cgggctggag gcgccgagca ccgagcgccg ccgggagtcg agcgccggcc
 721 gcggagctct tgcgaccccg ccaggacccg aacagagccc gggggcggcg ggccggagcc
 781 ggggacgcgg gcacacgccc gctcgcacaa gccacggcgg actctcccga ggcggaacct
 841 ccacgccgag cgagggtcag tttgaaaagg aggatcgagc tcactgtgga gtatccatgg
 901 agatgtggag ccttgtcacc aacctctaac tgcagaactg ggatgtggag ctggaagtgc
 961 ctcctcttct gggctgtgct ggtcacagcc acactctgca ccgctaggcc gtccccgacc
1021 ttgcctgaac aagcccagcc ctggggagcc cctgtggaag tggagtcctt cctggtccac
1081 cccggtgacc tgctgcagct tcgctgtcgg ctgcgggacg atgtgcagag catcaactgg
1141 ctgcgggacg gggtgcagct ggcggaaagc aaccgcaccc gcatcacagg ggaggaggtg
1201 gaggtgcagg actccgtgcc cgcagactcc ggcctctatg cttgcgtaac cagcagcccc
1261 tcgggcagtg acaccaccta cttctccgtc aatgtttcag atgctctccc ctcctcggag
1321 gatgatgatg atgatgatga ctcctcttca gaggagaaag aaacagataa caccaaacca
1381 aaccgtatgc ccgtagctcc atattggaca tccccagaaa agatggaaaa gaaattgcat
1441 gcagtgccgg ctgccaagac agtgaagttc aaatgccctt ccagtgggac cccaaacccc
1501 acactgcgct ggttgaaaaa tggcaaagaa ttcaaacctg accacagaat tggaggctac
1561 aaggtccgtt atgccacctg gagcatcata atggactctg tggtgccctc tgacaagggc
1621 aactacacct gcattgtgga gaatgagtac ggcagcatca ccacacata ccagctggat
1681 gtcgtggagc ggtcccctca ccggcccatc ctgcaagcag ggttgcccgc caacaaaaca
1741 gtggccctgg gtagcaacgt ggagttcatg tgtaaggtgt acagtgaccc gcagccgcac
1801 atccagtggc taaagcacat cgaggtgaat gggagcaaga ttggcccaga caacctgcct
1861 tatgtccaga tcttgaagac tgctggagtt aataccaccg acaaagagat ggaggtgctt
1921 cacttaagaa atgtctcctt tgaggacgca ggggagtata cgtgcttggc gggtaactct
1981 atcggactct cccatcactc tgcatggttg accgttctgg aagccctgga agaggccg
2041 gcagtgatga cctcgccccct gtacctggag atcatcatct attgcacagg ggccttcctc
2101 atctcctgca tggtggggtc ggtcatcgtc tacaagatga agagtggtac caagaagagt
2161 gacttccaca gccagatggc tgtgcacaag ctggccaaga gcatccctct gcgcagacag
2221 gtgtctgctg actccagtgc atccatgaac tctggggttc ttctggttcg gccatcacgg
2281 ctctcctcca gtgggactcc catgctagca ggggtctctg agtatgagct tcccgaagac
2341 cctcgctggg agctgcctcg ggacagactg gtcttaggca aacccctggg agagggctgc
2401 tttgggcagg tggtgttggc agaggctatc gggctggaca aggacaaacc caaccgtgtg
2461 accaaagtgg ctgtgaagat gttgaagtcg gacgcaacag agaaagactt gtcagacctg
2521 atctcagaaa tggagatgat gaagatgatc gggaagcata agaatatcat caacctgctg
2581 ggggcctgca cgcaggatgg tccttgtat gtcatcgtgg agtatgcctc caagggcaac
2641 ctgcgggagt acctgcaggc ccggaggccc ccagggctgg aatactgcta caaccccagc
2701 cacaacccag aggagcagct ctcctccaag gacctggtgt cctgcgccta ccaggtggcc
2761 cgaggcatgg agtatctggc ctccaagaag tgcatacacc gagacctggc agccaggaat
2821 gtcctggtga cagaggacaa tgtgatgaag atagcagact ttggcctcgc acgggacatt
2881 caccacatcg actactataa aaagacaacc aacggccgac tgcctgtgaa gtggatggca
```

2941 cccgaggcat tatttgaccg gatctacacc caccagagtg atgtgtggtc tttcggggtg
3001 ctcctgtggg agatcttcac tctgggcggc tccccatacc ccggtgtgcc tgtggaggaa
3061 cttttcaagc tgctgaagga gggtcaccgc atggacaagc ccagtaactg caccaacgag
3121 ctgtacatga tgatgcggga ctgctggcat gcagtgccct cacagagacc caccttc**aag**
3181 **cagctggtgg aagacctgga ccgcatcgtg gccttgacct ccaaccag**ga gtacctggac
3241 ctgtccatgc ccctggacca gtactccccc agctttcccg acacccggag ctctacgtgc
3301 tcctcagggg aggattccgt cttctctcat gagccgctgc ccgaggagcc ctgcctgccc
3361 cgacacccag cccagcttgc caatggcgga ctcaaacgcc gctgactgcc acccacacgc
3421 cctccccaga ctccaccgtc agctgtaacc ctcacccaca gcccctgctg ggcccaccac
3481 ctgtccgtcc ctgtcccctt tcctgctggc aggagccggc tgcctaccag gggccttcct
3541 gtgtggcctg ccttcacccc actcagctca cctctccctc cacctcctct ccacctgctg
3601 gtgagaggtg caaagaggca gatctttgct gccagccact tcatccccta ccagatgttg
3661 gaccaacacc cctccctgcc accaggcact gcctggaggg cagggagtgg gagccaatga
3721 acaggcatgc aagtgagagc ttcctgagct ttctcctgtc ggtttggtct gttttgcctt
3781 cacccataag cccctcgcac tctggtggca ggtgccttgt cctcagggct acagcagtag
3841 ggaggtcagt gcttcgtgcc tcgattgaag gtgacctctg ccccagatag gtggtgccag
3901 tggcttatta attccgatac tagtttgctt tgctgaccaa atgcctggta ccagaggatg
3961 gtgaggcgaa ggccaggttg ggggcagtgt tgtggccctg gggcccagcc ccaaactggg
4021 ggctctgtat atagctatga agaaaacaca aagtgtataa atctgagtat atatttacat
4081 gtctttttaa aagggtcgtt accagagatt tacccatcgg gtaagatgct cctggtggct
4141 gggaggcatc agttgctata tattaaaaac aaaaaagaaa aaaaaggaaa atgttttaa
4201 aaaggtcata tatttttgc tacttttgct gttttatttt tttaaattat gttctaaacc
4261 tattttcagt ttaggtccct caataaaaat tgctgctgct tcatttatct atgggctgta
4321 tgaaaagggt gggaatgtcc actggaaaga agggacaccc acgggccctg gggctaggtc
4381 tgtcccgagg gcaccgcatg ctccccggcgc aggttccttg taacctcttc ttcctaggtc
4441 ctgcacccag acctcacgac gcacctcctg cctctccgct gcttttggaa agtcagaaaa
4501 agaagatgtc tgcttcgagg gcaggaaccc catccatgca gtagaggcgc tgggcagaga
4561 gtcaaggccc agcagccatc gaccatggat ggtttcctcc aaggaaaccg gtggggttgg
4621 gctggggagg gggcacctac ctaggaatag ccacggggta gagctacagt gattaagagg
4681 aaagcaaggg cgcggttgct cacgcctgta atcccagcac tttgggacac cgaggtgggc
4741 agatcacttc aggtcaggag tttgagacca gcctggccaa cttagtgaaa ccccatctct
4801 actaaaaatg caaaaattat ccaggcatgg tggcacacgc ctgtaatccc agctccacag
4861 gaggctgagg cagaatccct tgaagctggg aggcggaggt tgcagtgagc cgagattgcg
4921 ccattgcact ccagcctggg caacagagaa aacaaaaagg aaaacaaatg atgaaggtct
4981 gcagaaactg aaacccagac atgtgtctgc cccctctatg tgggcatggt tttgccagtg
5041 cttctaagtg caggagaaca tgtcacctga ggctagtttt gcattcaggt ccctggcttc
5101 gttttcttgtt ggtatgcctc cccagatcgt ccttcctgta tccatgtgac cagactgtat
5161 ttgttgggac tgtcgcagat cttggcttct tacagttctt cctgtccaaa ctccatcctg
5221 tccctcagga acgggggaa aattctccga atgtttttgg ttttttggct gcttggaatt
5281 tacttctgcc acctgctggt catcactgtc ctcactaagt ggattctggc tccccccgtac
5341 ctcatggctc aaactaccac tcctcagtcg ctatattaaa gcttatattt tgctggatta
5401 ctgctaaata caaaagaaag ttcaatatgt tttcatttct gtagggaaaa tgggattgct
5461 gctttaaatt tctgagctag ggatttttg gcagctgcag tgttggcgac tattgtaaaa
5521 ttctctttgt ttctctctgt aaatagcacc tgctaacatt acaatttgta tttatgttta
5581 aagaaggcat catttggtga acagaactag gaaatgaatt tttagctctt aaaagcattt
5641 gctttgagac cgcacaggag tgtctttcct tgtaaaacag tgatgataat ttctgccttg
5701 gccctacctt gaagcaatgt tgtgtgaagg gatgaagaat ctaaaagtct tcataagtcc
5761 ttgggagagg tgctagaaaa atataaggca ctatcataat tacagtgatg tccttgctgt
5821 tactactcaa atcacccaca aatttcccca aagactgcgc tagctgtcaa ataaaagaca
5881 gtgaaattga cctga

[00134] SEQ ID NO: 185 is the FGFR1 Amino Acid Sequence for isoform 1, having Genebank Accession No. NP_075598 (822 aa):

1 MWSWKCLLFW AVLVTATLCT ARPSPTLPEQ AQPWGAPVEV ESFLVHPGDL LQLRCRLRDD
61 VQSINWLRDG VQLAESNRTR ITGEEVEVQD SVPADSGLYA CVTSSPSGSD TTYFSVNVSD
121 ALPSSEDDDD DDDSSSEEKE TDNTKPNRMP VAPYWTSPEK MEKKLHAVPA AKTVKFKCPS
181 SGTPNPTLRW LKNGKEFKPD HRIGGYKVRY ATWSIIMDSV VPSDKGNYTC IVENEYGSIN
241 HTYQLDVVER SPHRPILQAG LPANKTVALG SNVEFMCKVY SDPQPHIQWL KHIEVNGSKI

```
301 GPDNLPYVQI LKTAGVNTTD KEMEVLHLRN VSFEDAGEYT CLAGNSIGLS HHSAWLTVLE
361 ALEERPAVMT SPLYLEIIIY CTGAFLISCM VGSVIVYKMK SGTKKSDFHS QMAVHKLAKS
421 IPLRRQVTVS ADSSASMNSG VLLVRPSRLS SSGTPMLAGV SEYELPEDPR WELPRDRLVL
481 GKPLGEGCFG QVVLAEAIGL DKDKPNRVTK VAVKMLKSDA TEKDLSDLIS EMEMMKMIGK
541 HKNIINLLGA CTQDGPLYVI VEYASKGNLR EYLQARRPPG LEYCYNPSHN PEEQLSSKDL
601 VSCAYQVARG MEYLASKKCI HRDLAARNVL VTEDNVMKIA DFGLARDIHH IDYYKKTTNG
661 RLPVKWMAPE ALFDRIYTHQ SDVWSFGVLL WEIFTLGGSP YPGVPVEELF KLLKEGHRMD
721 KPSNCTNELY MMMRDCWHAV PSQRPTFKQL VEDLDRIVAL TSNQEYLDLS MPLDQYSPSF
781 PDTRSSTCSS GEDSVFSHEP LPEEPCLPRH PAQLANGGLK RR
```

[00135] SEQ ID NO: 186 is the FGFR1 Nucleotide Sequence for isoform 1, having Genebank Accession No. NM_023110 (5917 bp):

```
   1 agatgcaggg gcgcaaacgc caaaggagac caggctgtag gaagagaagg gcagagcgcc
  61 ggacagctcg gcccgctccc cgtcctttgg ggccgcggct ggggaactac aaggcccagc
 121 aggcagctgc aggggcggga ggcggaggag ggaccagcgc gggtgggagt gagagagcga
 181 gccctcgcgc cccgccggcg catagcgctc ggagcgctct tgcggccaca ggcgcggcgt
 241 cctcggcggc gggcggcagc tagcgggagc cgggacgccg gtgcagccgc agcgcgcgga
 301 ggaacccggg tgtgccggga gctgggcggc cacgtccgga cgggaccgag acccctcgta
 361 gcgcattgcg gcgacctcgc cttccccggc cgcgagcgcg ccgctgcttg aaaagccgcg
 421 gaacccaagg acttttctcc ggtccgagct cggggcgccc cgcagggcgc acggtacccg
 481 tgctgcagtc gggcacgccg cggcgccggg gcctccgcag ggcgatggag cccggtctgc
 541 aaggaaagtg aggcgccgcc gctgcgttct ggaggagggg ggcacaaggt ctggagaccc
 601 cgggtggcgg acgggagccc tcccccccgcc ccgcctccgg ggcaccagct ccggctccat
 661 tgttcccgcc cgggctggag gcgccgagca ccgagcgccg ccgggagtcg agcgccggcc
 721 gcggagctct tgcgaccccg ccaggacccg aacagagccc gggggcggcg ggccggagcc
 781 ggggacgcgg gcacacgccc gctcgcacaa gccacgggcgg actctcccga ggcggaacct
 841 ccacgccgag cgagggtcag tttgaaaagg aggatcgagc tcactgtgga gtatccatgg
 901 agatgtggag ccttgtcacc aacctctaac tgcagaactg ggatgtggag ctggaagtgc
 961 ctcctcttct gggctgtgct ggtcacagcc acactctgca ccgctaggcc gtccccgacc
1021 ttgcctgaac aagcccagcc ctggggagcc cctgtggaag tggagtcctt cctggtccac
1081 cccggtgacc tgctgcagct tcgctgtcgg ctgcgggacg atgtgcagag catcaactgg
1141 ctgcgggacg gggtgcagct ggcggaaagc aaccgcaccc gcatcacagg ggaggaggtg
1201 gaggtgcagg actccgtgcc cgcagactcc ggcctctatg cttgcgtaac cagcagcccc
1261 tcgggcagtg acaccaccta cttctccgtc aatgtttcag atgctctccc ctcctcggag
1321 gatgatgatg atgatgatga ctcctcttca gaggagaaag aaacagataa caccaaacca
1381 aaccgtatgc ccgtagctcc atattggaca tccccagaaa agatggaaaa gaaattgcat
1441 gcagtgccgg ctgccaagac agtgaagttc aaatgccctt ccagtgggac cccaaacccc
1501 acactgcgct ggttgaaaaa tggcaaagaa ttcaaacctg accacagaat tggaggctac
1561 aaggtccgtt atgccacctg gagcatcata atggactctg tggtgccctc tgacaagggc
1621 aactacacct gcattgtgga gaatgagtac ggcagcatca accacacata ccagctggat
1681 gtcgtggagc ggtcccctca ccggcccatc ctgcaagcag ggttgcccgc caacaaaaca
1741 gtggccctgg gtagcaacgt ggagttcatg tgtaaggtgt acagtgaccc gcagccgcac
1801 atccagtggc taaagcacat cgaggtgaat gggagcaaga ttggcccaga caacctgcct
1861 tatgtccaga tcttgaagac tgctggagtt aataccaccg acaaagagat ggaggtgctt
1921 cacttaagaa atgtctcctt tgaggacgca ggggagtata cgtgcttggc gggtaactct
1981 atcggactct cccatcactc tgcatggttg accgttctgg aagccctgga agagaggccg
2041 gcagtgatga cctcgcccct gtacctggag atcatcatct attgcacagg ggccttcctc
2101 atctcctgca tggtggggtc ggtcatcgtc tacaagatga agagtggtac caagaagagt
2161 gacttccaca gccagatggc tgtgcacaag ctggccaaga gcatccctct gcgcagacag
2221 gtaacagtgt ctgctgactc cagtgcatcc atgaactctg gggttcttct ggttcggcca
2281 tcacggctct cctccagtgg gactcccatg ctagcagggg tctctgagta tgagcttccc
2341 gaagaccctc gctgggagct gcctcgggac agactggtct taggcaaacc cctgggagag
2401 ggctgctttg ggcaggtggt gttggcagag gctatcgggc tggacaagga caaacccaac
2461 cgtgtgacca aagtggctgt gaagatgttg aagtcggacg caacagagaa agacttgtca
2521 gacctgatct cagaaatgga gatgatgaag atgatcggga agcataagaa tatcatcaac
2581 ctgctggggg cctgcacgca ggatggtccc ttgtatgtca tcgtggagta tgcctccaag
2641 ggcaacctgc gggagtacct gcaggcccgg aggccccccag ggctggaata ctgctacaac
```

2701 cccagccaca acccagagga gcagctctcc tccaaggacc tggtgtcctg cgcctaccag
2761 gtggccccgag gcatggagta tctggcctcc aagaagtgca tacaccgaga cctggcagcc
2821 aggaatgtcc tggtgacaga ggacaatgtg atgaagatag cagactttgg cctcgcacgg
2881 gacattcacc acatcgacta ctataaaaag acaaccaacg gccgactgcc tgtgaagtgg
2941 atggcacccg aggcattatt tgaccggatc tacacccacc agagtgatgt gtggtctttc
3001 ggggtgctcc tgtgggagat cttcactctg ggcggctccc catacccccgg tgtgcctgtg
3061 gaggaacttt tcaagctgct gaaggagggt caccgcatgg acaagcccag taactgcacc
3121 aacgagctgt acatgatgat gcgggactgc tggcatgcag tgccctcaca gagacccacc
3181 ttcaagcagc tggtggaaga cctggaccgc atcgtggcct tgacctccaa ccaggagtac
3241 ctggacctgt ccatgccccct ggaccagtac tcccccagct ttcccgacac ccggagctct
3301 acgtgctcct caggggagga ttccgtcttc tctcatgagc cgctgcccga ggagccctgc
3361 ctgccccgac acccagccca gcttgccaat ggcggactca aacgccgctg actgccaccc
3421 acacgccctc cccagactcc accgtcagct gtaaccctca cccacagccc ctgctgggcc
3481 caccacctgt ccgtccctgt ccccttttcct gctggcagga gccggctgcc taccaggggc
3541 cttcctgtgt ggcctgcctt caccccactc agctcacctc tccctccacc tcctctccac
3601 ctgctggtga gaggtgcaaa gaggcagatc tttgctgcca gccacttcat cccctcccag
3661 atgttggacc aacacccctc cctgccacca ggcactgcct ggagggcagg gagtgggagc
3721 caatgaacag gcatgcaagt gagagcttcc tgagctttct cctgtcggtt tggtctgttt
3781 tgccttcacc cataagcccc tcgcactctg gtggcaggtg ccttgtcctc agggctacag
3841 cagtagggag gtcagtgctt cgtgcctcga ttgaaggtga cctctgcccc agataggtgg
3901 tgccagtggc ttattaattc cgatactagt ttgctttgct gaccaaatgc ctggtaccag
3961 aggatggtga ggcgaaggcc aggttggggg cagtgttgtg gccctggggc ccagccccaa
4021 actgggggct ctgtatatag ctatgaagaa aacacaaagt gtataaatct gagtatatat
4081 ttacatgtct ttttaaaagg gtcgttacca gagatttacc catcgggtaa gatgctcctg
4141 gtggctggga ggcatcagtt gctatatatt aaaaacaaaa aagaaaaaaa aggaaaatgt
4201 ttttaaaaag gtcatatatt ttttgctact tttgctgttt tatttttta aattatgttc
4261 taaacctatt ttcagtttag gtccctcaat aaaaattgct gctgcttcat ttatctatgg
4321 gctgtatgaa aagggtggga atgtccactg gaaagaaggg acacccacgg gccctgggc
4381 taggtctgtc ccgagggcac cgcatgctcc cggcgcaggt tccttgtaac ctcttcttcc
4441 taggtcctgc acccagacct cacgacgcac ctcctgcctc tccgctgctt ttggaaagtc
4501 agaaaaagaa gatgtctgct tcgagggcag gaaccccatc catgcagtag aggcgctggg
4561 cagagagtca aggcccagca gccatcgacc atggatggtt tcctccaagg aaaccggtgg
4621 ggttgggctg gggaggggc acctacctag gaatagccac ggggtagagc tacagtgatt
4681 aagaggaaag caagggcgcg gttgctcacg cctgtaatcc cagcactttg ggacaccgag
4741 gtgggcagat cacttcaggt caggagtttg agaccagcct ggccaactta gtgaaacccc
4801 atctctacta aaaatgcaaa aattatccag gcatggtggc acacgcctgt aatcccagct
4861 ccacaggagg ctgaggcaga atcccttgaa gctgggaggc ggaggttgca gtgagccgag
4921 attgcgccat tgcactccag cctgggcaac agagaaaaca aaaaggaaaa caaatgatga
4981 aggtctgcag aaactgaaac ccagacatgt gtctgccccc tctatgtggg catggttttg
5041 ccagtgcttc taagtgcagg agaacatgtc acctgaggct agttttgcat tcaggtccct
5101 ggcttcgttt cttgttggta tgcctcccca gatcgtcctt cctgtatcca tgtgaccaga
5161 ctgtatttgt tgggactgtc gcagatcttg gctttcttaca gttcttcctg tccaaactcc
5221 atcctgtccc tcaggaacgg ggggaaaatt ctccgaatgt ttttggtttt ttggctgctt
5281 ggaatttact tctgccacct gctggtcatc actgtcctca ctaagtggat tctggctccc
5341 ccgtacctca tggctcaaac taccactcct cagtcgctat attaaagctt atattttgct
5401 ggattactgc taaatacaaa agaaagttca atatgttttc atttctgtag ggaaaatggg
5461 attgctgctt taaatttctg agctagggat tttttggcag ctgcagtgtt ggcgactatt
5521 gtaaaattct ctttgtttct ctctgtaaat agcacctgct aacattacaa tttgtatttta
5581 tgtttaaaga aggcatcatt tggtgaacag aactaggaaa tgaattttta gctcttaaaa
5641 gcatttgctt tgagaccgca caggagtgtc tttccttgta aaacagtgat gataatttct
5701 gccttggccc taccttgaag caatgttgtg tgaagggatg aagaatctaa aagtcttcat
5761 aagtcccttgg gagaggtgct agaaaaatat aaggcactat cataattaca gtgatgtcct
5821 tgctgttact actcaaatca cccacaaatt tccccaaaga ctgcgctagc tgtcaaataa
5881 aagacagtga aattgacctg aaaaaaaaaa aaaaaaa

[00136] The Genbank ID for the TACC1 gene is 6867. Three isoforms are listed for TACC1, e.g., having Genebank Accession Nos. NP_006274 (corresponding nucleotide sequence NM_001174063); NP_001167535 (corresponding nucleotide sequence

NM_001174064); NP_001167536 (corresponding nucleotide sequence NM_001174065).

[00137] SEQ ID NO: 148 is the TACC1 Amino Acid Sequence for isoform 1, having Genebank Accession No. NP_006274 (805 aa):

```
  1 MAFSPWQILS PVQWAKWTWS AVRGGAAGED EAGGPEGDPE EEDSQAETKS LSFSSDSEGN
 61 FETPEAETPI RSPFKESCDP SLGLAGPGAK SQESQEADEQ LVAEVVEKCS SKTCSKPSEN
121 EVPQQAIDSH SVKNFREEPE HDFSKISIVR PFSIETKDST DISAVLGTKA AHGCVTAVSG
181 KALPSSPPDA LQDEAMTEGS MGVTLEASAE ADLKAGNSCP ELVPSRRSKL RKPKPVPLRK
241 KAIGGEFSDT NAAVEGTPLP KASYHFSPEE LDENTSPLLG DAREQKSPPD LKETPGTLSS
301 DTNDSGVELG EESRSSPLKL EFDFTEDTGN IEARKALPRK LGRKLGSTLT PKIQKDGISK
361 SAGLEQPTDP VARDGPLSQT SSKPDPSQWE SPSFNPFGSH SVLQNSPPLS SEGSYHFDPD
421 NFDESMDPFK PTTTLTSSDF CSPTGNHVNE ILESPKKAKS RLITSGCKVK KHETQSLALD
481 ACSRDEGAVI SQISDISNRD GHATDEEKLA STSCGQKSAG AEVKGEPEED LEYFECSNVP
541 VSTINHAFSS SEAGIEKETC QKMEEDGSTV LGLLESSAEK APVSVSCGGE SPLDGICLSE
601 SDKTAVLTLI REEIITKEIE ANEWKKKYEE TRQEVLEMRK IVAEYEKTIA QMIEDEQRTS
661 MTSQKSFQQL TMEKEQALAD LNSVERSLSD LFRRYENLKG VLEGFKKNEE ALKKCAQDYL
721 ARVKQEEQRY QALKIHAEEK LDKANEEIAQ VRTKAKAESA ALHAGLRKEQ MKVESLERAL
781 QQKNQEIEEL TKICDELIAK LGKTD
```

[00138] SEQ ID NO: 149 is the TACC1 Nucleotide Sequence for isoform 1, having Genebank Accession No. NM_006283 (7802 bp):

```
   1 agctgatgcg cgccccgccg gccgggaggc gggagtccgc gagccgggag cgggagcagc
  61 agaggtctag cagccgggcg ccgcgggccg ggggcctgag gaggccacag gacgggcgtc
 121 ttcccggcta gtggagcccg gcgcgggggc cgctgcggcc gcaccgtgag gggaggaggc
 181 cgaggaggac gcagcgccgg ctgccggcgg gaggaagcgc tccaccaggg cccccgacgg
 241 cactcgttta accacatccg cgcctctgct ggaaacgctt gctggcgcct gtcaccggtt
 301 ccctccattt tgaaagggaa aaaggctctc cccacccatt cccctgcccc taggagctgg
 361 agccggagga gccgcgctca tggcgttcag cccgtggcag atcctgtccc ccgtgcagtg
 421 ggcgaaatgg acgtggtctg cggtacgcgg cggggccgcc ggcgaggacg aggctggcgg
 481 gccccgaggg cgaccccgagg aggaggattc gcaagccgag accaaatcct tgagtttcag
 541 ctcggattct gaaggtaatt ttgagactcc tgaagctgaa accccgatcc gatcaccttt
 601 caaggagtcc tgtgatccat cactcggatt ggcaggacct ggggccaaaa gccaagaatc
 661 acaagaagct gatgaacagc ttgtagcaga agtggttgaa aaatgttcat ctaagacttg
 721 ttctaaacct tcagaaaatg aagtgccaca gcaggccatt gactctcact cagtcaagaa
 781 tttcagagaa gaacctgaac atgattttag caaaatttcc atcgtgaggc cattttcaat
 841 agaaacgaag gattccacgg atatctcggc agtcctcgga acaaaagcag ctcatggctg
 901 tgtaactgca gtctcaggca aggctctgcc ttccagcccg ccagacgccc tccaggacga
 961 ggcgatgaca gaaggcagca tggggtcac cctcgaggcc tccgcagaag ctgatctaaa
1021 agctggcaac tcctgtccag agcttgtgcc cagcagaaga agcaagctga gaaagcccaa
1081 gcctgtcccc ctgaggaaga aagcaattgg aggagagttc tcagacacca acgctgctgt
1141 ggagggcaca cctctcccca aggcatccta tcacttcagt cctgaagagt tggatgagaa
1201 cacaagtcct ttgctaggag atgccaggtt ccagaagtct ccccctgacc ttaaagaaac
1261 tccccggcact ctcagtagtg acaccaacga ctcaggggtt gagctggggg aggagtcgag
1321 gagctcacct ctcaagcttg agtttgattt cacagaagat acaggaaaca tagaggccag
1381 gaaagcccct ccaaggaagc ttggcaggaa actgggtagc acactgactc ccaagataca
1441 aaaagatggc atcagtaagt cagcaggttt agaacagcct acagacccag tggcacgaga
1501 cgggcctctc tcccaaacat cttccaagcc agatcctagt cagtgggaaa gccccagctt
1561 caacccctttt gggagccact ctgttctgca gaactccccca cccctctctt ctgagggctc
1621 ctaccacttt gacccagata actttgacga atccatggat ccctttaaac caactacgac
1681 cttaacaagc agtgactttt gttctcccac tggtaatcac gttaatgaaa tcttagaatc
1741 acccaagaag gcaaagtcgc gtttaataac gagtggctgt aaggtgaaga agcatgaaac
1801 tcagtctctc gccctggatg catgttctcg ggatgaaggg gcagtgatct cccagatttc
1861 agacatttct aatagggatg gccatgctac tgatgaggag aaactggcat ccacgtcatg
1921 tggtcagaaa tcagctggtg ccgaggtgaa aggtgagcca gaggaagacc tggagtactt
```

1981 tgaatgttcc aatgttcctg tgtctaccat aaatcatgcg ttttcatcct cagaagcagg
2041 catagagaag gagacgtgcc agaagatgga agaagacggg tccactgtgc ttgggctgct
2101 ggagtcctct gcagagaagg cccctgtgtc ggtgtcctgt ggaggtgaga gccccctgga
2161 tgggatctgc ctcagcgaat cagacaagac agccgtgctc accttaataa gagaagagat
2221 aattactaaa gagattgaag caaatgaatg gaagaagaaa tacgaagaga cccggcaaga
2281 agttttggag atgaggaaaa ttgtagctga atatgaaaag actattgctc aaatgattga
2341 agatgaacaa aggacaagta tgacctctca gaagagcttc cagcaactga ccatggagaa
2401 ggaacaggcc ctggctgacc ttaactctgt ggaaaggtcc ctttctgatc tcttcaggag
2461 atatgagaac ctgaaaggtg ttctggaagg gttcaagaag aatgaagaag ccttgaagaa
2521 atgtgctcag gattacttag ccagagttaa acaagaggag cagcgatacc aggccctgaa
2581 aatccacgca gaagagaaac tggacaaagc caatgaagag attgctcagg ttcgaacaaa
2641 agcaaaggct gagagtgcag ctctccatgc tggactccgc aaagagcaga tgaaggtgga
2701 gtccctggaa agggccctgc agcagaagaa ccaagaaatt gaagaactga caaaaatctg
2761 tgatgagctg attgcaaagc tgggaaagac tgac*tgagac* actccccctg ttagctcaac
2821 agatctgcat ttggctgctt ctcttgtgac cacaattatc ttgccttatc caggaataat
2881 tgccccttg cagagaaaaa aaaaaactta aaaaaagcac atgcctactg ctgcctgtcc
2941 cgctttgctg ccaatgcaac agccctggaa gaaaccctag agggttgcat agtctagaaa
3001 ggagtgtgac ctgacagtgc tggagcctcc tagtttcccc ctatgaaggt tcccttaggc
3061 tgctgagttt gggtttgtga tttatcttta gtttgtttta aagtcatctt tactttccca
3121 aatgtgttaa atttgtaact cctctttggg gtcttctcca ccacctgtct gatttttttg
3181 tgatctgttt aatcttttaa ttttttagta tcagtggttt tatttaagga gacagtttgg
3241 cctattgtta cttccaattt ataatcaaga aggggctctg gatccccttt taaattacac
3301 acactctcac acacatacat gtatgtttat agatgctgct gctcttttcc ctgaagcata
3361 gtcaagtaag aactgctcta cagaaggaca tatttccttg gatgtgagac cctattttga
3421 aatagagtcc tgactcagaa caccaactta agaatttggg ggattaaaga tgtgaagacc
3481 acagtcttgg gttttcatat ctggagaaga ctatttgcca tgacgttttg ttgccctggt
3541 atttggacac tcctcagctt taatgggtgt ggccccttta gggttagtcc tcagactaat
3601 gatagtgtct gctttctgca tgaacggcaa tatgggactc cctccaagct agggtttggc
3661 aagtctgccc tagagtcatt tactctcctc tgcctccatt tgttaataca gaatcaacat
3721 ttagtcttca ttatcttttt ttttttttt gagacagagt ttcgatctat tttaagtatg
3781 tgaagaaaat ctacttgtaa aaggctcaga tcttaattaa aaggtaattg tagcacatta
3841 ccaattataa ggtgaagaaa tgttttttc ccaagtgtga tgcattgttc ttcagatgtt
3901 gaaaagaaag caaaaaatac cttctaactt aagacagaat ttttaacaaa atgagcagta
3961 aaagtcacat gaaccactcc aaaaatcagt gcattttgca tatttttaaa caaagacagc
4021 ttgttgaata ctgagaagag gagtgcaagg agaaggtctg tactaacaaa gccaaattcc
4081 tcaagctctt actggactca gttcagagtg gtgggccatt aaccccaaca tggaattttt
4141 ccatataaat ctcaatgaat tcccttttcat ttgaataggc aaacccaaat ccatgcaagt
4201 gttttaaagc actgtcctgt cttaatctta catgctgaaa gtcttcatgg tgatatgcac
4261 tatattcagt atacgtatgt tttcctactt ctcttgtaaa actgttgcat gatccaactt
4321 cagcaatgaa ttgtgcctag tggagaacct ctatagatct taaaaaatga attattcttt
4381 agcagtgtat tactcacatg ggtgcaatct ttagccccag ggaggtcaat aatgtctttt
4441 aaagccagaa gtcacatttt accaatatgc atttatcata attggtgctt aggctgtata
4501 ttcaagcctg ttgtcttaac attttgtata aaaaagaaca acagaaatta tctgtcattt
4561 gagaagtggc ttgacaatca tttgagcttt gaaagcagtc actgtggtgt aatatgaatg
4621 ctgtcctagt ggtcatagta ccaagggcac gtgtctcccc ttggtataac tgatttcctt
4681 tttagtcctc tactgctaaa taagttaatt ttgcattttg cagaaagaaa cattgattgc
4741 taaatctttt tgctgctgtg ttttggtgtt ttcatgttta cttgttttat attgatctgt
4801 tttaagtatg agaggcttat agtgccctcc attgtaaatc catagtcatc tttttaagct
4861 tattgtgttt aagaaagtag ctatgtgtta aacagaggtg atggcagccc ttccctagca
4921 cactggtgga agagacccct taagaacctg accccagtga atgaagctga tgcacaggga
4981 gcaccaaagg accttcgtta agtgataatt gtcctggcct ctcagccatg accgttatga
5041 ggaaatatcc cccattcgaa cttaacagat gcctcctctc caaagagaat taaaatcgta
5101 gcttgtacag atcaagagaa tatactgggc agaatgaagt atgtttgttt atttttcttt
5161 aaaaataaag gattttggaa ctctggagag taagaatata gtatagagtt tgcctcaaca
5221 catgtgaggg ccaaataacc tgctagctag gcagtaataa actctgttac agaagagaaa
5281 aagggccggg cacagtggct tattcctgta atcccaacac tgtggaaggc cgaggcagga
5341 ggatcacttg agtccaggag tttgaaacct acctaggcaa catggtgaaa ccttgtctct
5401 accaaaataa aaattagctg ggcatggtgg cacgtgcctg tggtcccagc tacttgggag
5461 gctgaggtgg gagcctggga ggtcaaggct gcagtgagcc atgatcatgc cactgcactc
5521 catcctgggt gacagcaaga tcttgtctca aaaaaaaaaa aaaaaaaaaa aaaaccagga
5581 gtgaaaaagg aaagtagaag gcagctgctg gcctagatgt tggtttggga atattaggtg

```
5641 atcctgttga gattctggat ccagagcaat ttctttagct tttgactttg ccaaagtgta
5701 gatagcctttt atccagcagt attttaagtg gggaatgcaa cgtgaggcca actgaacaat
5761 tccccccgtg gctgcccaga tagtcacagt caaggttgga gagtctcctt ccagccagtg
5821 acctacccaa acctttttgtt ctgtaaaact gctctggaaa taccgggaag cccagttttc
5881 tcacgtggtt tctagcttct tcagactcag cccaaattag gaagtgcaga agcacatgat
5941 ggtgaaaaac ctaggatttg gcagccttcc agaatggtat ggaatctgag ggaagattta
6001 tgtttcgttt tggaggatag ctcaagttga attttctttc cagccagtta ccctttcaac
6061 ctacccatac tttgtacaac tcttacacaa atacttagat atttattaga tagccctgaa
6121 ttcactctaa ttataaacag ggagtgtaaa ctgcccccag atgttcctgg gctgggtaaa
6181 agcagctgga gtgaagcact cattttccat aaaggtaaca aagggcagct cagtggttac
6241 tcaagctcaa aagggttttt ttaagagcaa gcattggtta agtctgtgta tactgagttg
6301 gaagtgattt cagcacattc ttttttagtg gagtgaaagt tctgaagccc ccttttaact
6361 tcctcttggt ttttcattat aattggtagc catctcatga actgtctctg actgttgtct
6421 ctttgtggtc atgtgattgt gagcttgctt tctgacttgc atttctgact ttatcctgtt
6481 gttaggaaga tagaaactag gttttgaaag attacatgat tcaagcgagg gattttaaag
6541 taaagatgta tttattctga agaatctaaa agataacaga ttatttgctt atgaaagaac
6601 aatatagtct gggaatccca gaatgtcaag ccaaaggtct aagaagtcat ctccttcaaa
6661 tactttaata aagaagtatt tcgaggagat atctgtccaa aaaggtttga ctggcctcca
6721 gattccagtt atttttaaaa agcaacttac cactaaatcc ttgagtctcc atagagtaac
6781 agtaaagaaa ctgatgtaac agactctcct ctcaaaggat ctcctctgga agagactatc
6841 agcggcagca ttctccaggg aagacccatc ccctagtgcc agagcttgca tcctggagac
6901 taaagattgc actttttttgt agtttttttgt ccaaatgcaa tcccatttct gtgcctctta
6961 gcatgcagtt agatttggac aaacaagatt cctaaggaat gactttatta actataatat
7021 ggttacagct attatataaa tatatattct ggttatagtt ctaatatgga gatgttgtgt
7081 gcaatgctgg cctgtggtgg tctgtgtaat gctttaactt gtatggagga ggccaggctc
7141 agagctgaga tgtggcctga accttccctg tatcgatcct ttaatttaga actgtcaaga
7201 tgtcactttc tccccctctg ccttttagtg gtatctgaca tatactcaaa acagtaattt
7261 cctggtcaca tcattaactg ctaattctgt atttataaag aattttcaga tggacatgta
7321 caaatttgaa ctcaaaccat ccccagtcca gatacagggc agcgtgtagg tgaccacacc
7381 agagcctcag cctcggtcct tctcagccgt cgggatagga tccaggcatt tcttttaaat
7441 ctcagaggta gcagtaaact tttcagtatt gctgttagca agtgtgtgtt tgccaataga
7501 tacccattat actaatgtgc caagtaaatg ttcattgcac atctgcttcc actgtgttcc
7561 cacgggtgcc atgaagtgtg tgaggagccc ctcatctgga gggatgagtg ctgcgttgac
7621 tactgctatc aggattgtgt tgtgtggaat attcatctac ataaatttta tatgcacagt
7681 aatttccctt tttatatgtc aagtaactat ttgtaaaagt tatactcaca aattattata
7741 atgattacta atatattttt tccatgtttc attgcctgaa taaaaactgt ttaccactgt
7801 ta
```

[00139] SEQ ID NO: 150 is the amino acid sequence of the FGFR1-TACC1 fusion protein.

```
MWSWKCLLFWAVLVTATLCTARPSPTLPEQDALPSSEDDDDDDDSSSEEKETDNTKPNPVAPYWTSPEKMEKKLH
AVPAAKTVKFKCPSSGTPNPTLRWLKNGKEFKPDHRIGGYKVRYATWSIIMDSVVPSDKGNYTCIVENEYGSINH
TYQLDVVERSPHRPILQAGLPANKTVALGSNVEFMCKVYSDPQPHIQWLKHIEVNGSKIGPDNLPYVQILKTAGV
NTTDKEMEVLHLRNVSFEDAGEYTCLAGNSIGLSHHSAWLTVLEALEERPAVMTSPLYLEIIIYCTGAFLISCMV
GSVIVYKMKSGTKKSDFHSQMAVHKLAKSIPLRRQVTVSADSSASMNSGVLLVRPSRLSSSGTPMLAGVSEYELP
EDPRWELPRDRLVLGKPLGEGCFGQVVLAEAIGLDKDKPNRVTKVAVKMLKSDATEKDLSDLISEMEMMKMIGKH
KNIINLLGACTQDGPLYVIVEYASKGNLREYLQARRPPGLEYCYNPSHNPEEQLSSKDLVSCAYQVARGMEYLAS
KKCIHRDLAARNVLVTEDNVMKIADFGLARDIHHIDYYKKTTNGRLPVKWMAPEALFDRIYTHQSDVWSFGVLLW
EIFTLGGSPYPGVPVEELFKLLKEGHRMDKPSNCTNELYMMMRDCWHAVPSQRPTFKQLVEDLDRIVALTSNQ**GL**
**LESSAEKA**PVSVSCGGESPLDGICLSESDKTAVLTLIREEIITKEIEANEWKKKYEETRQEVLEMRKIVAEYEKT
IAQMIEDEQRTSMTSQKSFQQLTMEKEQALADLNSVERSLSDLFRRYENLKGVLEGFKKNEEALKKCAQDYLARV
KQEEQRYQALKIHAEEKLDKANEEIAQVRTKAKAESAALHAGLRKEQMKVESLERALQQKNQEIEELTKICDELI
AKLGKTD
```

[00140] SEQ ID NO: 151 is the nucleotide sequence that encodes the FGFR1-TACC1

fusion protein.

atgtggagctggaagtgcctcctcttctgggctgtgctggtcacagccacactctgcaccgctaggccgtccccg
accttgcctgaacaagcccagccctggggagcccctgtggaagtggagtccttcctggtccaccccggtgacctg
ctgcagcttcgctgtcggctgcgggacgatgtgcagagcatcaactggctgcgggacggggtgcagctggcggaa
agcaaccgcacccgcatcacaggggaggaggtggaggtgcaggactccgtgcccgcagactccggcctctatgct
tgcgtaaccagcagccccteggcagtgacaccacctacttctccgtcaatgtttcagatgctctcccctectcg
gaggatgatgatgatgatgactcctcttcagaggagaaagaaacagataacaccaaaccaaaccgtatgccc
gtagctccatattggacatccccagaaaagatggaaaagaaattgcatgcagtgccggctgccaagacagtgaag
ttcaaatgcccttccagtgggacccaaaccccacactgcgctggttgaaaaatggcaaagaattcaaacctgac
cacagaattggaggctacaaggtccgttatgccacctggagcatcataatggactctgtggtgccctctgacaag
ggcaactacacctgcattgtggagaatgagtacggcagcatcaaccacacataccagctggatgtcgtggagcgg
tcccctcaccggcccatcctgcaagcagggttgcccgccaacaaaacagtggccctgggtagcaacgtggagttc
atgtgtaaggtgtacagtgacccgcagccgcacatccagtggctaaagcacatcgaggtgaatgggagcaagatt
ggcccagacaacctgccttatgtccagatcttgaagactgctggagttaataccaccgacaaagagatggaggtg
cttcacttaagaaatgtctcctttgaggacgcagggagtatacgtgcttggcgggtaactctatcggactctcc
catcactctgcatggttgaccgttctggaagccctggaagagaggccggcagtgatgacctcgcccctgtacctg
gagatcatcatctattgcacagggccttcctcatctcctgcatggtggggtcggtcatcgtctacaagatgaag
agtggtaccaagaagagtgacttccacagccagatggctgtgcacaagctggccaagagcatccctctgcgcaga
caggtgtctgctgactccagtgcatccatgaactctgggqttcttctggttcggccatcacggctctcctccagt
gggactcccatgctagcaggggtctctgagtatgagcttcccgaagaccctcgctgggagctgcctcgggacaga
ctggtcttaggcaaacccctgggagagggctgctttgggcaggtggtgttggcagaggctatcgggctggacaag
gacaaacccaaccgtgtgaccaaagtggctgtgaagatgttgaagtcggacgcaacagagaaagacttgtcagac
ctgatctcagaaatggagatgatgaagatgatcgggaagcataagaatatcatcaacctgctgggggcctgcacg
caggatggtcccttgtatgtcatcgtggagtatgcctccaagggcaacctgcgggagtacctgcaggcccggagg
ccccccagggctggaatactgctacaaccccagccacaacccagaggagcagctctcctccaaggacctggtgtcc
tgcgcctaccaggtggcccgaggcatggagtatctggcctccaagaagtgcatacaccgagacctggcagccagg
aatgtcctggtgacagaggacaatgtgatgaagatagcagactttggcctcgcacgggacattcaccacatcgac
tactataaaaagacaaccaacggccgactgcctgtgaagtggatggcacccgaggcattatttgaccggatctac
acccaccagagtgatgtgtggtctttcggggtgctcctgtgggagatcttcactctgggcggctccccataccc
ggtgtgcctgtggaggaacttttcaagctgctgaaggagggtcaccgcatggacaagcccagtaactgcaccaac
gagctgtacatgatgatgcgggactgctggcatgcagtgccctcacagagacccaccttcaagcagctggtggaa
gacctggaccgcatcgtggccttgacctccaaccagtgggctgctggagtcctctgcagagaaggccccctgtgtc
ggtgtcctgtggaggtgagagccccctggatgggatctgcctcagcgaatcagacaagacagccgtgctcacctt
aataagagaagagataattactaaagagattgaagcaaatgaatggaagaagaaatacgaagagacccggcaaga
agttttggagatgaggaaaattgtagctgaatatgaaaagactattgctcaaatgattgaagatgaacaaaggac
aagtatgacctctcagaagagcttccagcaactgaccatggagaaggaacaggccctggctgaccttaactctgt
ggaaaggtccctttctgatctcttcaggagatatgagaacctgaaaggtgttctggaagggttcaagaagaatga
agaagccttgaagaaatgtgctcaggattacttagccagagttaaacaagaggagcagcgataccaggccctgaa
aatccacgcagaagagaaactggacaaagccaatgaagagattgctcaggttcgaacaaaagcaaaggctgagag
tgcagctctccatgctggactccgcaaagagcagatgaaggtggagtccctggaaagggccctgcagcagaagaa
ccaagaaattgaagaactgacaaaaatctgtgatgagctgattgcaaagctgggaaagactgac

[00141] The Genbank ID for the FGFR2 gene is 2263. Eight isoforms are listed for FGFR2, e.g., having Genebank Accession Nos. NP_000132 (corresponding nucleotide sequence NM_000141); NP_001138385 (corresponding nucleotide sequence NM_001144913); NP_001138386 (corresponding nucleotide sequence NM_001144914); NP_001138387 (corresponding nucleotide sequence NM_001144915); NP_001138388 (corresponding nucleotide sequence NM_001144916); NP_001138389 (corresponding nucleotide sequence NM_001144917); NP_001138390 (corresponding nucleotide sequence NM_001144918); NP_001138391 (corresponding nucleotide sequence NM_001144919); NP_075259 (corresponding nucleotide sequence NM_022970).

[00142] SEQ ID NO: 152 is the FGFR2 Amino Acid Sequence for isoform 1, having Genebank Accession No. NP_000132 (821 aa):

```
  1 MVSWGRFICL VVVTMATLSL ARPSFSLVED TTLEPEEPPT KYQISQPEVY VAAPGESLEV
 61 RCLLKDAAVI SWTKDGVHLG PNNRTVLIGE YLQIKGATPR DSGLYACTAS RTVDSETWYF
121 MVNVTDAISS GDDEDDTDGA EDFVSENSNN KRAPYWTNTE KMEKRLHAVP AANTVKFRCP
181 AGGNPMPTMR WLKNGKEFKQ EHRIGGYKVR NQHWSLIMES VVPSDKGNYT CVVENEYGSI
241 NHTYHLDVVE RSPHRPILQA GLPANASTVV GGDVEFVCKV YSDAQPHIQW IKHVEKNGSK
301 YGPDGLPYLK VLKAAGVNTT DKEIEVLYIR NVTFEDAGEY TCLAGNSIGI SFHSAWLTVL
361 PAPGREKEIT ASPDYLEIAI YCIGVFLIAC MVVTVILCRM KNTTKKPDFS SQPAVHKLTK
421 RIPLRRQVTV SAESSSSMNS NTPLVRITTR LSSTADTPML AGVSEYELPE DPKWEFPRDK
481 LTLGKPLGEG CFGQVVMAEA VGIDKDKPKE AVTVAVKMLK DDATEKDLSD LVSEMEMMKM
541 IGKHKNIINL LGACTQDGPL YVIVEYASKG NLREYLRARR PPGMEYSYDI NRVPEEQMTF
601 KDLVSCTYQL ARGMEYLASQ KCIHRDLAAR NVLVTENNVM KIADFGLARD INNIDYYKKT
661 TNGRLPVKWM APEALFDRVY THQSDVWSFG VLMWEIFTLG GSPYPGIPVE ELFKLLKEGH
721 RMDKPANCTN ELYMMMRDCW HAVPSQRPTF KQLVEDLDRI LTLTTNEEYL DLSQPLEQYS
781 PSYPDTRSSC SSGDDSVFSP DPMPYEPCLP QYPHINGSVK T
```

[00143] SEQ ID NO: 153 is the FGFR2 Nucleotide Sequence for isoform 1, having Genebank Accession No. NM_000141 (4654 bp):

```
   1 ggcggcggct ggaggagagc gcggtggaga gccgagcggg cgggcggcgg gtgcggagcg
  61 ggcgagggag cgcgcgcggc cgccacaaag ctcgggcgcc gcggggctgc atgcggcgta
 121 cctggcccgg cgcggcgact gctctccggg ctggcggggg ccggccgcga gccccggggg
 181 ccccgaggcc gcagcttgcc tgcgcgctct gagccttcgc aactcgcgag caaagtttgg
 241 tggaggcaac gccaagcctg agtcctttct tcctctcgtt ccccaaatcc gagggcagcc
 301 cgcgggcgtc atgcccgcgc tcctccgcag cctggggtac gcgtgaagcc cgggaggctt
 361 ggcgccggcg aagacccaag gaccactctt ctgcgtttgg agttgctccc cgcaacccgg
 421 ggctcgtcgc tttctccatc ccgacccacg cggggcgcgg ggacaacaca ggtcgcggag
 481 gagcgttgcc attcaagtga ctgcagcagc agcggcagcg cctcggttcc tgagcccacc
 541 gcaggctgaa ggcattgcgc gtagtccatg cccgtagagg aagtgtgcag atgggattaa
 601 cgtccacatg gagatatgga agaggaccgg ggattggtac cgtaaccatg gtcagctggg
 661 gtcgtttcat ctgcctggtc gtggtcacca tggcaacctt gtccctggcc cggccctcct
 721 tcagtttagt tgaggatacc acattagagc cagaagagcc accaaccaaa taccaaatct
 781 ctcaaccaga agtgtacgtg gctgcgccag gggagtcgct agaggtgcgc tgcctgttga
 841 aagatgccgc cgtgatcagt tggactaagg atggggtgca cttggggccc aacaatagga
 901 cagtgcttat tggggagtac ttgcagataa agggcgccac gcctagagac tccggcctct
 961 atgcttgtac tgccagtagg actgtagaca gtgaaacttg gtacttcatg gtgaatgtca
1021 cagatgccat ctcatccgga gatgatgagg atgacaccga tggtgcggaa gattttgtca
1081 gtgagaacag taacaacaag agagcaccat actggaccaa cacagaaaag atggaaaagc
1141 ggctccatgc tgtgcctgcg gccaacactg tcaagtttcg ctgcccagcc ggggggaacc
1201 caatgccaac catgcggtgg ctgaaaaacg ggaaggagtt taagcaggag catcgcattg
1261 gaggctacaa ggtacgaaac cagcactgga gcctcattat ggaaagtgtg gtcccatctg
1321 acaagggaaa ttatacctgt gtagtggaga atgaatacgg gtccatcaat cacacgtacc
1381 acctggatgt tgtggagcga tcgcctcacc ggcccatcct ccaagccgga ctgccggcaa
1441 atgcctccac agtggtcgga ggagacgtag agtttgtctg caaggtttac agtgatgccc
1501 agccccacat ccagtggatc aagcacgtgg aaaagaacgg cagtaaatac gggccccgacg
1561 ggctgcccta cctcaaggtt ctcaaggccg ccggtgttaa caccacggac aaagagattg
1621 aggttctcta tattcggaat gtaacttttg aggacgctgg ggaatatacg tgcttggcgg
1681 gtaattctat tgggatatcc tttcactctg catggttgac agttctgcca gcgcctggaa
1741 gagaaaagga gattacagct tccccagact acctggagat agccatttac tgcataggga
1801 tcttcttaat cgcctgtatg gtggtaacag tcatcctgtg ccgaatgaag aacacgacca
1861 agaagccaga cttcagcagc cagccggctg tgcacaagct gaccaaacgt atccccctgc
1921 ggagacaggt aacagtttcg gctgagtcca gctcctccat gaactccaac accccgctgg
1981 tgaggataac aacacgcctc tcttcaacgg cagacacccc catgctggca ggggtctccg
2041 agtatgaact tccagaggac ccaaaatggg agtttccaag agataagctg acactgggca
2101 agccccctggg agaaggttgc tttgggcaag tggtcatggc ggaagcagtg ggaattgaca
```

2161 aagacaagcc caaggaggcg gtcaccgtgg ccgtgaagat gttgaaagat gatgccacag
2221 agaaagacct ttctgatctg gtgtcagaga tggagatgat gaagatgatt gggaaacaca
2281 agaatatcat aaatcttctt ggagcctgca cacaggatgg gcctctctat gtcatagttg
2341 agtatgcctc taaaggcaac ctccgagaat acctccgagc ccggaggcca cccgggatgg
2401 agtactccta tgacattaac cgtgttcctg aggagcagat gaccttcaag gacttggtgt
2461 catgcaccta ccagctggcc agaggcatgg agtacttggc ttcccaaaaa tgtattcatc
2521 gagatttagc agccagaaat gttttggtaa cagaaaacaa tgtgatgaaa atagcagact
2581 ttggactcgc cagagatatc aacaatatag actattacaa aaagaccacc aatgggcggc
2641 ttccagtcaa gtggatggct ccagaagccc tgtttgatag agtatacact catcagagtg
2701 atgtctggtc cttcggggtg ttaatgtggg agatcttcac tttagggggc tcgccctacc
2761 cagggattcc cgtggaggaa ctttttaagc tgctgaagga aggacacaga atggataagc
2821 cagccaactg caccaacgaa ctgtacatga tgatgaggga ctgttggcat gcagtgccct
2881 cccagagacc aacgttcaag cagttggtag aagacttgga tcgaattctc actctcacaa
2941 ccaatgagga atacttggac ctcagccaac ctctcgaaca gtattcacct agttaccctg
3001 acacaagaag ttcttgttct tcaggagatg attctgtttt ttctccagac cccatgcctt
3061 acgaaccatg ccttcctcag tatccacaca taaacggcag tgttaaaaca tgaatgactg
3121 tgtctgcctg tccccaaaca ggacagcact gggaacctag ctacactgag cagggagacc
3181 atgcctccca gagcttgttg tctccacttg tatatatgga tcagaggagt aaataattgg
3241 aaaagtaatc agcatatgtg taaagattta tacagttgaa aacttgtaat cttccccagg
3301 aggagaagaa ggtttctgga gcagtggact gccacaagcc accatgtaac ccctctcacc
3361 tgccgtgcgt actggctgtg gaccagtagg actcaaggtg gacgtgcgtt ctgccttcct
3421 tgttaatttt gtaataattg gagaagattt atgtcagcac acacttacag agcacaaatg
3481 cagtatatag gtgctggatg tatgtaaata tattcaaatt atgtataaat atatattata
3541 tatttacaag gagttatttt ttgtattgat tttaaatgga tgtcccaatg cacctagaaa
3601 attggtctct ctttttttaa tagctatttg ctaaatgctg ttcttacaca taatttctta
3661 attttcaccg agcagaggtg gaaaaatact tttgctttca gggaaaatgg tataacgtta
3721 atttattaat aaattggtaa tatacaaaac aattaatcat ttatagtttt ttttgtaatt
3781 taagtggcat ttctatgcag gcagcacagc agactagtta atctattgct tggacttaac
3841 tagttatcag atcctttgaa aagagaatat ttacaatata tgactaattt ggggaaaatg
3901 aagttttgat ttatttgtgt ttaaatgctg ctgtcagacg attgttctta gacctcctaa
3961 atgccccata ttaaaagaac tcattcatag gaaggtgttt cattttggtg tgcaaccctg
4021 tcattacgtc aacgcaacgt ctaactggac ttcccaagat aaatggtacc agcgtcctct
4081 taaaagatgc cttaatccat tccttgagga cagaccttag ttgaaatgat agcagaatgt
4141 gcttctctct ggcagctggc cttctgcttc tgagttgcac attaatcaga ttagcctgta
4201 ttctcttcag tgaattttga taatggcttc cagactcttt ggcgttggag acgcctgtta
4261 ggatcttcaa gtcccatcat agaaaattga aacacagagt tgttctgctg atagttttgg
4321 ggatacgtcc atctttttaa gggattgctt tcatctaatt ctggcaggac ctcaccaaaa
4381 gatccagcct catacctaca tcagacaaaa tatcgccgtt gttccttctg tactaaagta
4441 ttgtgttttg ctttggaaac acccactcac tttgcaatag ccgtgcaaga tgaatgcaga
4501 ttacactgat cttatgtgtt acaaaattgg agaaagtatt taataaaacc tgttaatttt
4561 tatactgaca ataaaaatgt ttctacagat attaatgtta acaagacaaa ataaatgtca
4621 cgcaacttat ttttttaata aaaaaaaaaa aaaa

[00144] The Genbank ID for the TACC2 gene is 10579. Four isoforms are listed for TACC2, e.g., having Genebank Accession Nos. NP_996744 (corresponding nucleotide sequence NM_206862); NP_996743 (corresponding nucleotide sequence NM_206861); NP_996742 (corresponding nucleotide sequence NM_206860); NP_008928 (corresponding nucleotide sequence NM_006997).

[00145] SEQ ID NO: 154 is the TACC2 Amino Acid Sequence for isoform a, having Genebank Accession No. NP_996744 (2948 aa):

1 MGNENSTSDN QRTLSAQTPR SAQPPGNSQN IKRKQQDTPG SPDHRDASSI GSVGLGGFCT

```
  61 ASESSASLDP CLVSPEVTEP RKDPQGARGP EGSLLPSPPP SQEREHPSSS MPFAECPPEG
 121 CLASPAAAPE DGPQTQSPRR EPAPNAPGDI AAAFPAERDS STPYQEIAAV PSAGRERQPK
 181 EEGQKSSFSF SSGIDQSPGM SPVPLREPMK APLCGEGDQP GGFESQEKEA AGGFPPAESR
 241 QGVASVQVTP EAPAAAQQGT ESSAVLEKSP LKPMAPIPQD PAPRASDRER GQGEAPPQYL
 301 TDDLEFLRAC HLPRSNSGAA PEAEVNAASQ ESCQQPVGAY LPHAELPWGL PSPALVPEAG
 361 GSGKEALDTI DVQGHPQTGM RGTKPNQVVC VAAGGQPEGG LPVSPEPSLL TPTEEAHPAS
 421 SLASFPAAQI PIAVEEPGSS SRESVSKAGM PVSADAAKEV VDAGLVGLER QVSDLGSKGE
 481 HPEGDPGEVP APSPQERGEH LNTEQSHEVQ PGVPPPPLPK EQSHEVQPGA PPPPLPKAPS
 541 ESARGPPGPT DGAKVHEDST SPAVAKEGSR SPGDSPGGKE EAPEPPDGGD PGNLQGEDSQ
 601 AFSSKRDPEV GKDELSKPSS DAESRDHPSS HSAQPPRKGG AGHTDGPHSQ TAEADASGLP
 661 HKLGEEDPVL PPVPDGAGEP TVPEGAIWEG SGLQPKCPDT LQSREGLGRM ESFLTLESEK
 721 SDFPPTPVAE VAPKAQEGES TLEIRKMGSC DGEGLLTSPD QPRGPACDAS RQEFHAGVPH
 781 PPQGENLAAD LGLTALILDQ DQQGIPSCPG EGWIRGAASE WPLLSSEKHL QPSQAQPETS
 841 IFDVLKEQAQ PPENGKETSP SHPGFKDQGA DSSQIHVPVE PQEDNNLPTH GGQEQALGSE
 901 LQSQLPKGTL SDTPTSSPTD MVWESSLTEE SELSAPTRQK LPALGEKRPE GACGDGQSSR
 961 VSPPAADVLK DFSLAGNFSR KETCCTGQGP NKSQQALADA LEEGSQHEEA CQRHPGASEA
1021 ADGCSPLWGL SKREMASGNT GEAPPCQPDS VALLDAVPCL PALAPASPGV TPTQDAPETE
1081 ACDETQEGRQ QPVPAPQQKM ECWATSDAES PKLLASFPSA GEQGGEAGAA ETGGSAGAGD
1141 PGKQQAPEKP GEATLSCGLL QTEHCLTSGE EASTSALRES CQAEHPMASC QDALLPAREL
1201 GGIPRSTMDF STHQAVPDPK ELLLSGPPEV AAPDTPYLHV DSAAQRGAED SGVKAVSSAD
1261 PRAPGESPCP VGEPPLALEN AASLKLFAGS LAPLLQPGAA GGEIPAVQAS SGSPKARTTE
1321 GPVDSMPCLD RMPLLAKGKQ ATGEEKAATA PGAGAKASGE GMAGDAAGET EGSMERMGEP
1381 SQDPKQGTSG GVDTSSEQIA TLTGFPDFRE HIAKIFEKPV LGALATPGEK AGAGRSAVGK
1441 DLTRPLGPEK LLDGPPGVDV TLLPAPPARL QVEKKQQLAG EAEISHLALQ DPASDKLLGP
1501 AGLTWERNLP GAGVGKEMAG VPPTLREDER PEGPGAAWPG LEGQAYSQLE RSRQELASGL
1561 PSPAATQELP VERAAAFQVA PHSHGEEAVA QDRIPSGKQH QETSACDSPH GEDGPGDFAH
1621 TGVPGHVPRS TCAPSPQREV LTVPEANSEP WTLDTLGGER RPGVTAGILE MRNALGNQST
1681 PAPPTGEVAD TPLEPGKVAG AAGEAEGDIT LSTAETQACA SGDLPEAGTT RTFSVVAGDL
1741 VLPGSCQDPA CSDKAPGMEG TAALHGDSPA RPQQAKEQPG PERPIPAGDG KVCVSSPPEP
1801 DETHDPKLQH LAPEELHTDR ESPRPGPSML PSVPKKDAPR VMDKVTSDET RGAEGTESSP
1861 VADDIIQPAA PADLESPTLA ASSYHGDVVG QVSTDLIAQS ISPAAAHAGL PPSAAEHIVS
1921 PSAPAGDRVE ASTPSCPDPA KDLSRSSDSE EAFETPESTT PVKAPPAPPP PPPEVIPEPE
1981 VSTQPPPEEP GCGSETVPVP DGPRSDSVEG SPFRPPSHSF SAVFDEDKPI ASSGTYNLDF
2041 DNIELVDTFQ TLEPRASDAK NQEGKVNTRR KSTDSVPISK STLSRSLSLQ ASDFDGASSS
2101 GNPEAVALAP DAYSTGSSSA SSTLKRTKKP RPPSLKKKQT TKKPTETPPV KETQQEPDEE
2161 SLVPSGENLA SETKTESAKT EGPSPALLEE TPLEPAVGPK AACPLDSESA EGVVPPASGG
2221 GRVQNSPPVG RKTLPLTTAP EAGEVTPSDS GGQEDSPAKG LSVRLEFDYS EDKSSWDNQQ
2281 ENPPPTKKIG KKPVAKMPLR RPKMKKTPEK LDNTPASPPR SPAEPNDIPI AKGTYTFDID
2341 KWDDPNFNPF SSTSKMQESP KLPQQSYNFD PDTCDESVDP FKTSSKTPSS PSKSPASFEI
2401 PASAMEANGV DGDGLNKPAK KKKTPLKTDT FRVKKSPKRS PLSDPPSQDP TPAATPETPP
2461 VISAVVHATD EEKLAVTNQK WTCMTVDLEA DKQDYPQPSD LSTFVNETKF SSPTEELDYR
2521 NSYEIEYMEK IGSSLPQDDD APKKQALYLM FDTSQESPVK SSPVRMSESP TPCSGSSFEE
2581 TEALVNTAAK NQHPVPRGLA PNQESHLQVP EKSSQKELEA MGLGTPSEAI EITAPEGSFA
2641 SADALLSRLA HPVSLCGALD YLEPDLAEKN PPLFAQKLQE ELEFAIMRIE ALKLARQIAL
2701 ASRSHQDAKR EAAHPTDVSI SKTALYSRIG TAEVEKPAGL LFQQPDLDSA LQIARAEIIT
2761 KEREVSEWKD KYEESRREVM EMRKIVAEYE KTIAQMIEDE QREKSVSHQT VQQLVLEKEQ
2821 ALADLNSVEK SLADLFRRYE KMKEVLEGFR KNEEVLKRCA QEYLSRVKKE EQRYQALKVH
2881 AEEKLDRANA EIAQVRGKAQ QEQAAHQASL RKEQLRVDAL ERTLEQKNKE IEELTKICDE
2941 LIAKMGKS
```

[00146] SEQ ID NO: 155 is the TACC2 Nucleotide Sequence for isoform a, having Genebank Accession No. NM_206862 (9706 bp)

```
   1 gcctgctcca agggaaggat caggagagaa gaaacgcaaa tcccagaacc gtgccaacat
  61 ataaaacccc acattaaggg ttgtacagtg cactgggatt tctcaagtca cccgcttggt
 121 cctcttccaa gtatacttta cttcctttca ttcctctcta aaacttttt aaaaactttc
 181 actcctgctc taaaagttat cttggtttct tactctacct tatgcccctt gggcgaattt
 241 tttcctctga ggagggaaga atagagttgc tgctgcagac acatcagatt ccctactggt
```

```
301 aacagctgga gtgcgtcacc tctgacaaaa ttctggggac gctgggaaca ctgaatcaac
361 atgggcaatg agaacagcac ctcggacaac cagaggactt tatcagctca gactccaagg
421 tccgcgcagc cacccgggaa cagtcagaat ataaaaagga agcagcagga cacgcccgga
481 agccctgacc acagagacgc gtccagcatt ggcagcgttg ggcttggagg cttctgcacc
541 gcttctgaga gttctgccag cctggatcca tgccttgtgt ccccagaggt gactgagcca
601 aggaaggacc cacagggagc cagggggcca gaaggttctt tgctgcccag cccaccaccg
661 tcccaggagc gagagcaccc ctcgtcctcc atgccctttg ccgagtgtcc cccggaaggt
721 tgcttggcaa gtccagcagc ggcacctgaa gatggtcctc agactcagtc tcccaggagg
781 gaacctgccc caaatgcccc aggagacatc gcggcggcat ttcccgctga gagggacagc
841 tctactccat accaagagat tgctgccgtc cccagtgctg gaagagagag acagccgaag
901 gaagaaggac agaagtcctc cttctccttc tccagtggca tcgaccagtc acctggaatg
961 tcgccagtac ccctcagaga gccaatgaag gcaccgctgt gtggagaggg ggaccagcct
1021 ggtggttttg agtcccaaga gaaagaggct gcaggtggct ttccccctgc agagtccagg
1081 caggggtgg cttctgtgca agtgacccct gaggcccctg ctgcagccca gcagggcaca
1141 gaaagctcag cggtcttgga gaagtccccc ctaaaaccca tggccccgat cccacaagat
1201 ccagccccaa gagcctcaga cagagaaaga ggccaagggg aggcgccgcc tcagtattta
1261 acagatgact tggaattcct cagggcctgc catctcccta ggagcaattc aggggctgcc
1321 ccagaagcag aagtgaatgc cgcttcccag gagagctgcc agcagccagt gggagcatat
1381 ctgccgcacg cagagctgcc ctggggcttg ccaagtcctg ccctggtgcc agaggctggg
1441 ggctctggga aggaggctct ggacaccatt gatgttcagg gtcacccaca gacagggatg
1501 cgaggaacca agcccaatca agttgtctgt gtggcagcag gcggccagcc cgaaggggt
1561 ttgcctgtga gccctgaacc ttccctgctc actccgactg aggaagcaca tccagcttca
1621 agcctcgctt cattcccagc tgctcagatt cctattgctg tagaagaacc tggatcatca
1681 tccagggaat cagtttccaa ggctgggatg ccagtttctg cagatgcagc caaagaggtg
1741 gtggatgcag ggttggtggg actggagagg caggtgtcag atcttggaag caagggagag
1801 catccagaag gggaccctgg agaggttcct gccccatcac cccaggagag gggagagcac
1861 ttgaacacgg agcaaagcca tgaggtccaa ccaggagtac cacccctcc tcttcccaag
1921 gagcaaagcc atgaggtcca accaggagca ccacccctc ctcttcccaa ggcaccaagt
1981 gaaagtgcca gagggccacc ggggccaacg gatggagcca aggtccatga agattccaca
2041 agcccagccg tggctaaaga aggaagcaga tcacctggtg acagccctgg aggaaaggag
2101 gaagccccag agccacctga tggtggagac ccagggaacc tgcaaggaga ggactctcag
2161 gctttcagca gcaagcgtga tccagaagta ggcaaagatg agctttcaaa gccaagcagt
2221 gatgcagaga gcagagacca tcccagctca cactcagcac agccacccag aaagggggt
2281 gctgggcaca cggacgggcc ccactctcag acagcagagg ctgatgcatc tggcctacca
2341 cacaagctgg gtgaggagga ccccgtcctg ccccctgtgc cagatggagc tggtgagccc
2401 actgttcccg aaggagccat ctgggagggg tcaggattgc agcccaaatg tcctgacacc
2461 cttcagagca gggaaggatt gggaagaatg gagtctttcc tgactttaga atcagagaaa
2521 tcagattttc caccaactcc tgttgcagag gttgcaccca aagcccagga aggtgagagc
2581 acattggaaa taaggaagat gggcagctgt gatggggagg gcttgctgac gtccccagat
2641 caaccccgcg ggccggcgtg tgatgcgtcg agacaggaat ttcatgctgg ggtgccacat
2701 ccccccccagg gggagaactt ggcagcagac ctggggctca cggcactcat cctggaccaa
2761 gatcagcagg gaatcccatc ctgcccaggg gaaggctgga taagaggagc tgcatccgag
2821 tggcccctac tatcttctga gaagcatctc cagccatccc aggcacaacc agagacatcc
2881 atctttgacg tgctcaagga gcaggcccag ccacctgaaa atgggaaaga gacttctcca
2941 agccatccag gttttaagga ccagggagca gattcttccc aaatccatgt acctgtggaa
3001 cctcaggaag ataacaactt gcccactcat ggaggacagg agcaggcttt gggatcagaa
3061 cttcaaagtc agctccccaa aggcaccctg tctgatactc caacttcatc tcccactgac
3121 atggtttggg agagttctct gacagaagag tcagaattgt cagcaccaac gagacagaag
3181 ttgcctgcac taggggagaa gcggccagag ggagcatgcg gtgatggtca gtcctcgagg
3241 gtctcgcctc cagcagcaga tgtcttaaaa gacttttctc ttgcagggaa cttcagcaga
3301 aaggaaactt gctgcactgg gcagggggcca aacaagtctc aacaggcatt ggctgatgcc
3361 ttggaagaag gcagccagca tgaagaagca tgtcaaaggc atccaggagc ttctgaagca
3421 gctgatggtt gttccccact ctggggcttg agtaagaggg agatggcaag tggaaacaca
3481 ggggaggccc caccttgtca gcctgactca gtagctctcc tggatgcagt tccctgcctg
3541 ccagccctgg cgcccgccag ccccggagtc acacccaccc aggatgcccc agagacagag
3601 gcatgtgatg aaacccagga aggcaggcag caaccagtgc cggccccgca gcagaaaatg
3661 gagtgctggg ccacttcgga tgcagagtcc ccaaagcttc ttgcaagttt cccatcagct
3721 ggggagcaag gtggtgaagc cggggctgct gagactggtg gcagcgctgg tgcaggagac
3781 ccaggaaagc agcaggctcc ggagaaacct ggagaagcta ctttgagttg tggcctcctt
3841 cagactgagc actgccttac ctccggggag gaagcttcta cctctgccct acgtgagtcc
3901 tgccaagctg agcaccccat ggccagctgc caggatgcct tgctgccagc cagagagctg
```

3961 ggtgggattc ccaggagcac catggatttt tctacacacc aggctgtccc agacccaaag
4021 gagctcctgc tgtctgggcc accagaagtg gctgctcctg acacccctta cctgcatgtc
4081 gacagtgctg cccagagagg agcagaagac agtggagtga aagctgtttc ctctgcagac
4141 cccagagctc ctggcgaaag ccctgtcct gtaggggagc ccccacttgc cttggaaaat
4201 gctgcctcct tgaagctgtt tgctggctcc ctcgcccccc tgttgcaacc aggagctgca
4261 ggtggggaaa tccctgcagt gcaagccagc agtggtagtc ccaaagccag aaccactgag
4321 ggaccagtgg actccatgcc atgcctggac cggatgccac ttctggccaa gggcaagcag
4381 gcaacagggg aagagaaagc agcaacagct ccaggtgcag gtgccaaggc cagtggggag
4441 ggcatggcag gtgatgcagc aggagagaca gagggcagca tggagaggat gggagagcct
4501 tcccaggacc caaagcaggg cacatcaggt ggtgtggaca caagctctga gcaaatcgcc
4561 accctcactg gcttcccaga cttcagggag cacatcgcca agatcttcga gaagcctgtg
4621 ctcggagccc tggccacacc tggagaaaag gcaggagctg ggaggagtgc agtgggtaaa
4681 gacctcacca ggccattggg cccagagaag cttctagatg ggcctccagg agtggatgtc
4741 acccttctcc ctgcacctcc tgctcgactc caggtggaga agaagcaaca gttggctgga
4801 gaggctgaga tttcccatct ggctctgcaa gatccagctt cagacaagct tctgggtcca
4861 gcagggctga cctgggagcg gaacttgcca ggtgccggtg tggggaagga gatggcaggt
4921 gtcccaccca cactgaggga agacgagagg ccagaggggc ctggggcagc ctggccaggc
4981 ctggaaggcc aggcttactc acagctggag aggagcaggc aggaattagc ttcaggtctt
5041 ccttcaccag cagctactca ggagctccct gtggagagag ctgctgcctt ccaggtggct
5101 ccccatagcc atggagaaga ggccgtggcc caagacagaa ttccttctgg aaagcagcac
5161 caggaaacat ctgcctgcga cagtccacat ggagaagatg gtcccgggga ctttgctcac
5221 acaggggttc caggacatgt gccaaggtcc acgtgtgccc cttctcctca gagggaggtt
5281 ttgactgtgc ctgaggccaa cagtgagccc tggacccttg acacgcttgg gggtgaaagg
5341 agacccggag tcactgctgg catcttggaa atgcgaaatg ccctgggcaa ccagagcacc
5401 cctgcaccac caactggaga agtggcagac actcccctgg agcctggcaa ggtggcaggc
5461 gctgctgggg aagcagaggg tgacatcacc ctgagcacag ctgagacaca ggcatgtgcg
5521 tccggtgatc tgcctgaagc aggtactacg aggacattct ccgttgtggc aggtgacttg
5581 gtgctgccag gaagctgtca ggacccagcc tgctctgaca aggctccggg gatggagggt
5641 acagctgccc ttcatgggga cagcccagcc aggccccagc aggctaagga gcagccaggg
5701 cctgagcgcc ccattccagc tggggatggg aaggtgtgcg tctcctcacc tccagagcct
5761 gacgaaactc acgacccgaa gctgcaacat ttggctccag aagagctcca cactgacaga
5821 gagagcccca ggcctggccc atccatgtta ccttcggttc ctaagaagga tgctccaaga
5881 gtcatggata aagtcacttc agatgagacc agaggtgcgg aaggaacaga aagttcacct
5941 gtggcagatg atatcatcca gcccgctgcc cccgcagacc tggaaagccc aaccttagct
6001 gcctcttcct accacggtga tgttgttggc caggtctcta cggatctgat agcccagagc
6061 atctccccag ctgctgccca tgcgggtctt cctccctcgg ctgcagaaca catagtttcg
6121 ccatctgccc cagctggtga cagagtagaa gcttccactc cctcctgccc agatccggcc
6181 aaggacctca gcaggagttc cgattctgaa gaggcatttg agaccccgga gtcaacgacc
6241 cctgtcaaag ctccgccagc tccaccccca ccacccccgg aagtcatccc agaacccgag
6301 gtcagcacac agccaccccc ggaagaacca ggatgtggtt ctgagacagt ccctgtccct
6361 gatggcccac ggagcgactc ggtggaagga agtcccttcc gtcccccgtc acactccttc
6421 tctgccgtct tcgatgaaga caagccgata gccagcagtg ggacttacaa cttggacttt
6481 gacaacattg agcttgtgga taccttttcag accttggagc ctcgtgcctc agacgctaag
6541 aatcaggagg gcaaagtgaa cacacggagg aagtccacgg attccgtccc catctctaag
6601 tctacactgt cccggtcgct cagcctgcaa gccagtgact ttgatggtgc ttcttcctca
6661 ggcaatcccg aggccgtggc ccttgcccca gatgcatata gcacgggttc cagcagtgct
6721 tctagtaccc ttaagcgaac taaaaaaccg aggccgcctt ccttaaaaaa gaaacagacc
6781 accaagaaac ccacagagac ccccccagtg aaggagacgc aacaggagcc agatgaagag
6841 agccttgtcc ccagtgggga gaatctagca tctgagacga aaacggaatc tgccaagacg
6901 gaaggtccta gcccagcctt attggaggag acgccccttg agcccgctgt ggggcccaaa
6961 gctgcctgcc ctctggactc agagagtgca gaagggttg tcccccggc ttctggaggt
7021 ggcagagtgc agaactcacc ccctgtcggg aggaaaacgc tgcctcttac cacggccccg
7081 gaggcagggg aggtaacccc atcggatagc gggggcaag aggactctcc agccaaaggg
7141 ctctccgtaa ggctggagtt tgactattct gaggacaaga gtagttggga caaccagcag
7201 gaaaaccccc ctcctaccaa aaagataggc aaaaagccag ttgccaaaat gcccctgagg
7261 aggccaaaga tgaaaaagac acccgagaaa cttgacaaca ctcctgcctc acctcccaga
7321 tcccctgctg aacccaatga catccccatt gctaaaggta cttacacctt tgatattgac
7381 aagtgggatg accccaattt taacccttt tcttccacct caaaaatgca ggagtctccc
7441 aaactgcccc aacaatcata caactttgac ccagacacct gtgatgagtc cgttgacccc
7501 tttaagacat cctctaagac ccccagctca ccttctaaat ccccagcctc ctttgagatc
7561 ccagccagtg ctatggaagc caatggagtg gacggggatg ggctaaacaa gcccgccaag

7621 aagaagaaga cgcccctaaa gactgacaca tttagggtga aaaagtcgcc aaaacggtct
7681 cctctctctg atccaccttc ccaggacccc accccagctg ctacaccaga aacaccacca
7741 gtgatctctg cggtggtcca cgccacagat gaggaaaagc tggcggtcac caaccagaag
7801 tggacgtgca tgacagtgga cctagaggct gacaaacagg actacccgca gccctcggac
7861 ctgtccacct ttgtaaacga gaccaaattc agttcaccca ctgaggagtt ggattacaga
7921 aactcctatg aaattgaata tatggagaaa attggctcct ccttacctca ggacgacgat
7981 gccccgaaga agcaggcctt gtaccttatg tttgacactt ctcaggagag ccctgtcaag
8041 tcatctcccg tccgcatgtc agagtccccg acgccgtgtt cagggtcaag ttttgaagag
8101 actgaagccc ttgtgaacac tgctgcgaaa aaccagcatc ctgtcccacg aggactggcc
8161 cctaaccaag agtcacactt gcaggtgcca gagaaatcct cccagaagga gctggaggcc
8221 atgggcttgg gcacccctc agaagcgatt gaaattacag ctcccgaggg ctcctttgcc
8281 tctgctgacg ccctcctcag caggctagct caccccgtct ctctctgtgg tgcacttgac
8341 tatctggagc ccgacttagc agaaaagaac cccccactat tcgctcagaa actccaggag
8401 gagttagagt ttgccatcat gcggatagaa gccctgaagc tggccaggca gattgctttg
8461 gcttcccgca gccaccagga tgccaagaga gaggctgctc acccaacaga cgtctccatc
8521 tccaaaacag ccttgtactc ccgcatcggg accgctgagg tggagaaacc tgcaggcctt
8581 ctgttccagc agcccgacct ggactctgcc ctccagatcg ccagagcaga gatcataacc
8641 aaggagagag aggtctcaga atggaaagat aaatatgaag aaagcaggcg ggaagtgatg
8701 gaaatgagga aaatagtggc cgagtatgag aagaccatcg ctcagatgat agaggacgaa
8761 cagagagaga agtcagtctc ccaccagacg gtgcagcagc tggttctgga gaaggagcaa
8821 gccctggccg acctgaactc cgtggagaag tctctggccg acctcttcag aagatatgag
8881 aagatgaagg aggtcctaga aggcttccgc aagaatgaag aggtgttgaa gagatgtgcg
8941 caggagtacc tgtcccgggt gaagaaggag gagcagaggt accaggccct gaaggtgcac
9001 gcggaggaga aactggacag ggccaatgct gagattgctc aggttcgagg caaggcccag
9061 caggagcaag ccgcccacca ggccagcctg cggaaggagc agctgcgagt ggacgccctg
9121 gaaaggacgc tggagcagaa gaataaagaa atagaagaac tcaccaagat ttgtgacgaa
9181 ctgattgcca aaatggggaa aagctaactc tgaaccgaat gttttggact taactgttgc
9241 gtgcaatatg accgtcggca cactgctgtt cctccagttc catggacagg ttctgttttc
9301 acttttcgt atgcactact gtatttcctt tctaaataaa attgatttga ttgtatgcag
9361 tactaaggag actatcagaa tttcttgcta ttggtttgca ttttcctagt ataattcata
9421 gcaagttgac ctcagagttc ctgtatcagg gagattgtct gattctctaa taaaagacac
9481 attgctgacc ttggccttgc cctttgtaca caagttccca gggtgagcag cttttggatt
9541 taatatgaac atgtacagcg tgcataggga ctcttgcctt aaggagtgta aacttgatct
9601 gcatttgctg atttgttttt aaaaaaacaa gaaatgcatg tttcaaataa aattctctat
9661 tgtaaataaa atttttttctt tggatcttgg caaaaaaaaa aaaaaa

[00147] SEQ ID NO: 158 is the amino acid sequence of the FGFR3-TACC3-1 fusion protein. The bolded text corresponds to the FGF3 protein:

MGAPACALALCVAVAIVAGASSESLGTEQRVVGRAAEVPGPEPGQQEQLVFGSGDAVELSCPPPGGGPMGPTVWV
KDGTGLVPSERVLVGPQRLQVLNASHEDSGAYSCRQRLTQRVLCHFSVRVTDAPSSGDDEDGEDEAEDTGVDTGA
PYWTRPERMDKKLLAVPAANTVRFRCPAAGNPTPSISWLKNGREFRGEHRIGGIKLRHQQWSLVMESVVPSDRGN
YTCVVENKFGSIRQTYTLDVLERSPHRPILQAGLPANQTAVLGSDVEFHCKVYSDAQPHIQWLKHVEVNGSKVGP
DGTPYVTVLKTAGANTTDKELEVLSLHNVTFEDAGEYTCLAGNSIGFSHHSAWLVVLPAEEELVEADEAGSVYAG
ILSYGVGFFLFILVVAAVTLCRLRSPPKKGLGSPTVHKISRFPLKRQVSLESNASMSSNTPLVRIARLSSGEGPT
LANVSELELPADPKWELSRARLTLGKPLGEGCFGQVVMAEAIGIDKDRAAKPVTVAVKMLKDDATDKDLSDLVSE
MEMMKMIGKHKNIINLLGACTQGGPLYVLVEYAAKGNLREFLRARRPPGLDYSFDTCKPPEEQLTFKDLVSCAYQ
VARGMEYLASQKCIHRDLAARNVLVTEDNVMKIADFGLARDVHNLDYYKKTTNGRLPVKWMAPEALFDRVYTHQS
DVWSFGVLLWEIFTLGGSPYPGIPVEELFKLLKEGHRMDKPANCTHDLYMIMRECWHAAPSQRPTFKQLVEDLDR
VLTVTSTDFKESALRKQSLYLKFDPLLRDSPGRPVPVATETSSMHGANETPSGRPREAKLVEFDFLGALDIPVPG
PPPGVPAPGGPPLSTGPIVDLLQYSQKDLDAVVKATQEENRELRSRCEELHGKNLELGKIMDRFEEVVYQAMEEV
QKQKELSKAEIQKVLKEKDQLTTDLNSMEKSFSDLFKRFEKQKEVIEGYRKNEESLKKCVEDYLARITQEGQRYQ
ALKAHAEEKLQLANEEIAQVRSKAQAEALALQASLRKEQMRIQSLEKTVEQKTKENEELTRICDDLISKMEKI

[00148] SEQ ID NO: 159 is the amino acid sequence of the FGFR3-TACC3-2 fusion protein. The bolded text corresponds to the FGF3 protein:

MGAPACALALCVAVAIVAGASSESLGTEQRVVGRAAEVPGPEPGQQEQLVFGSGDAVELSCPPPGGGPMGPTVWV
KDGTGLVPSERVLVGPQRLQVLNASHEDSGAYSCRQRLTQRVLCHFSVRVTDAPSSGDDEDGEDEAEDTGVDTGA
PYWTRPERMDKKLLAVPAANTVRFRCPAAGNPTPSISWLKNGREFRGEHRIGGIKLRHQQWSLVMESVVPSDRGN
YTCVVENKFGSIRQTYTLDVLERSPHRPILQAGLPANQTAVLGSDVEFHCKVYSDAQPHIQWLKHVEVNGSKVGP
DGTPYVTVLKTAGANTTDKELEVLSLHNVTFEDAGEYTCLAGNSIGFSHHSAWLVVLPAEEELVEADEAGSVYAG
ILSYGVGFFLFILVVAAVTLCRLRSPPKKGLGSPTVHKISRFPLKRQVSLESNASMSSNTPLVRIARLSSGEGPT
LANVSELELPADPKWELSRARLTLGKPLGEGCFGQVVMAEAIGIDKDRAAKPVTVAVKMLKDDATDKDLSDLVSE
MEMMKMIGKHKNIINLLGACTQGGPLYVLVEYAAKGNLREFLRARRPPGLDYSFDTCKPPEEQLTFKDLVSCAYQ
VARGMEYLASQKCIHRDLAARNVLVTEDNVMKIADFGLARDVHNLDYYKKTTNGRLPVKWMAPEALFDRVYTHQS
DVWSFGVLLWEIFTLGGSPYPGIPVEELFKLLKEGHRMDKPANCTHDLYMIMRECWHAAPSQRPTFKQLVEDLDR
VLTVTSTDVSAGSGLVPPAYAPPPAVPGHPSGRPREAKLVEFDFLGALDIPVPGPPPGVPAPGGPPLSTGPIVDL
LQYSQKDLDAVVKATQEENRELRSRCEELHGKNLELGKIMDRFEEVVYQAMEEVQKQKELSKAEIQKVLKEKDQL
TTDLNSMEKSFSDLFKRFEKQKEVIEGYRKNEESLKKCVEDYLARITQEGQRYQALKAHAEEKLQLANEEIAQVR
SKAQAEALALQASLRKEQMRIQSLEKTVEQKTKENEELTRICDDLISKMEKI

[00149] SEQ ID NO: 160 is the amino acid sequence of the FGFR3-TACC3-3 fusion protein. The bolded text corresponds to the FGF3 protein:

**MGAPACALALCVAVAIVAGASSESLGTEQRVVGRAAEVPGPEPGQQEQLVFGSGDAVELSCPPPGGGPMGPTVWV
KDGTGLVPSERVLVGPQRLQVLNASHEDSGAYSCRQRLTQRVLCHFSVRVTDAPSSGDDEDGEDEAEDTGVDTGA
PYWTRPERMDKKLLAVPAANTVRFRCPAAGNPTPSISWLKNGREFRGEHRIGGIKLRHQQWSLVMESVVPSDRGN
YTCVVENKFGSIRQTYTLDVLERSPHRPILQAGLPANQTAVLGSDVEFHCKVYSDAQPHIQWLKHVEVNGSKVGP
DGTPYVTVLKTAGANTTDKELEVLSLHNVTFEDAGEYTCLAGNSIGFSHHSAWLVVLPAEEELVEADEAGSVYAG
ILSYGVGFFLFILVVAAVTLCRLRSPPKKGLGSPTVHKISRFPLKRQVSLESNASMSSNTPLVRIARLSSGEGPT
LANVSELELPADPKWELSRARLTLGKPLGEGCFGQVVMAEAIGIDKDRAAKPVTVAVKMLKDDATDKDLSDLVSE
MEMMKMIGKHKNIINLLGACTQGGPLYVLVEYAAKGNLREFLRARRPPGLDYSFDTCKPPEEQLTFKDLVSCAYQ
VARGMEYLASQKCIHRDLAARNVLVTEDNVMKIADFGLARDVHNLDYYKKTTNGRLPVKWMAPEALFDRVYTHQS
DVWSFGVLLWEIFTLGGSPYPGIPVEELFKLLKEGHRMDKPANCTHDLYMIMRECWHAAPSQRPTFKQLVEDLDR
VLTVTSTD**VPGPPPGVPAPGGPPLSTGPIVDLLQYSQKDLDAVVKATQEENRELRSRCEELHGKNLELGKIMDRF
EEVVYQAMEEVQKQKELSKAEIQKVLKEKDQLTTDLNSMEKSFSDLFKRFEKQKEVIEGYRKNEESLKKCVEDYL
ARITQEGQRYQALKAHAEEKLQLANEEIAQVRSKAQAEALALQASLRKEQMRIQSLEKTVEQKTKENEELTRICD
DLISKMEKI

[00150] SEQ ID NO: 161 is the amino acid sequence of the FGFR3-TACC3-4 fusion protein. The bolded text corresponds to the FGF3 protein:

**MGAPACALALCVAVAIVAGASSESLGTEQRVVGRAAEVPGPEPGQQEQLVFGSGDAVELSCPPPGGGPMGPTVWV
KDGTGLVPSERVLVGPQRLQVLNASHEDSGAYSCRQRLTQRVLCHFSVRVTDAPSSGDDEDGEDEAEDTGVDTGA
PYWTRPERMDKKLLAVPAANTVRFRCPAAGNPTPSISWLKNGREFRGEHRIGGIKLRHQQWSLVMESVVPSDRGN
YTCVVENKFGSIRQTYTLDVLERSPHRPILQAGLPANQTAVLGSDVEFHCKVYSDAQPHIQWLKHVEVNGSKVGP
DGTPYVTVLKTAGANTTDKELEVLSLHNVTFEDAGEYTCLAGNSIGFSHHSAWLVVLPAEEELVEADEAGSVYAG
ILSYGVGFFLFILVVAAVTLCRLRSPPKKGLGSPTVHKISRFPLKRQVSLESNASMSSNTPLVRIARLSSGEGPT
LANVSELELPADPKWELSRARLTLGKPLGEGCFGQVVMAEAIGIDKDRAAKPVTVAVKMLKDDATDKDLSDLVSE
MEMMKMIGKHKNIINLLGACTQGGPLYVLVEYAAKGNLREFLRARRPPGLDYSFDTCKPPEEQLTFKDLVSCAYQ
VARGMEYLASQKCIHRDLAARNVLVTEDNVMKIADFGLARDVHNLDYYKKTTNGRLPVKWMAPEALFDRVYTHQS
DVWSFGVLLWEIFTLGGSPYPGIPVEELFKLLKEGHRMDKPANCTHDLYMIMRECWHAAPSQRPTFKQLVEDLDR
VLTVTSTDVKATQEENRELRSRCEELHGKNLELGKIMDRFEEVVYQAMEEVQKQKELSKAEIQKVLKEKDQLTTD
LNSMEKSFSDLFKRFEKQKEVIEGYRKNEESLKKCVEDYLARITQEGQRYQALKAHAEEKLQLANEEIAQVRSKA
QAEALALQASLRKEQMRIQSLEKTVEQKTKENEELTRICDDLISKMEKI

[00151] The Genbank ID for the FGFR4 gene is 2264. Three isoforms are listed for FGFR4, e.g., having Genebank Accession Nos. NP_002002 (corresponding nucleotide sequence NM_002011); NP_075252 (corresponding nucleotide sequence NM_022963);

NP_998812 (corresponding nucleotide sequence NM_213647).

[00152] As used herein, a "FGFR fusion molecule" can be a nucleic acid (e.g., synthetic, purified, and/or recombinant) which encodes a polypeptide corresponding to a fusion protein comprising a tyrosine kinase domain of an FGFR protein fused to a polypeptide that constitutively activates the tyrosine kinase domain of the FGFR protein, or a nucleic acid encoding a fusion protein comprising a transforming acidic coiled-coil (TACC) domain fused to a polypeptide with a tyrosine kinase domain, wherein the TACC domain constitutively activates the tyrosine kinase domain. It can also be a fusion protein comprising a tyrosine kinase domain of an FGFR protein fused to a polypeptide that constitutively activates the tyrosine kinase domain of the FGFR protein, or a fusion protein comprising a transforming acidic coiled-coil (TACC) domain fused to a polypeptide with a tyrosine kinase domain, wherein the TACC domain constitutively activates the tyrosine kinase domain. For example, a FGFR fusion molecule can include a FGFR1-TACC1 (e.g., comprising the amino acid sequence shown in SEQ ID NO: 150, or comprising the nucleic acid sequence shown in SEQ ID NO: 88), FGFR2-TACC2, FGFR3-TACC3 (e.g., comprising the amino acid sequence shown in SEQ ID NOS: 79, 158-160, or 161, or comprising the nucleic acid sequence shown in SEQ ID NOS: 80-82, 84, 94-144, or 145), or other FGFR-TACC fusion proteins (e.g., an N-terminal fragment of FGFR4 containing its tyrosine kinase domain fused to a fragment of TACC1, TACC2, or TACC3). For example, a FGFR fusion molecule can include a FGFR1-containing fusion comprising the amino acid sequence corresponding to Genebank Accession no. NP_001167534, NP_001167535, NP_001167536, NP_001167537, NP_001167538, NP_056934, NP_075593, NP_075594, or NP_075598; or a FGFR1-containing fusion comprising the nucleotide sequence corresponding to Genebank Accession no. NM_001174063, NM_001174064, NM_001174065, NM_001174066, NM_001174067, NM_015850, NM_023105, NM_023106, or NM_023110. For example, a FGFR fusion molecule can include a FGFR2-containing fusion comprising the amino acid sequence corresponding to Genebank Accession no. NP_000132, NP_001138385, NP_001138386, NP_001138387, NP_001138388, NP_001138389, NP_001138390, NP_001138391, or NP_075259; or a FGFR2-containing fusion comprising the nucleotide sequence corresponding to Genebank Accession no. NM_000141, NM_001144913, NM_001144914, NM_001144915, NM_001144916, NM_001144917, NM_001144918, NM_001144919, or NM_022970. For example, a FGFR fusion molecule can include a FGFR3-containing fusion comprising the amino acid sequence corresponding to Genebank Accession no. NP_000133,

NP_001156685, or NP_075254; or a FGFR3-containing fusion comprising the nucleotide sequence corresponding to Genebank Accession no. NM_000142, NM_001163213, or NM_022965. For example, a FGFR fusion molecule can include a FGFR4-containing fusion comprising the amino acid sequence corresponding to Genebank Accession no. NP_002002, NP_075252, or NP_998812; or a FGFR4-containing fusion comprising the nucleotide sequence corresponding to Genebank Accession no. NM_002011, NM_022963, or NM_213647. A FGFR fusion molecule can also include a tyrosine kinase domain of an FGFR protein fused to a protein encoded by any one of the genes listed in Table 7. A FGFR fusion molecule can include a variant of the above described examples, such as a fragment thereof.

[00153] **Table 7.** Fusion Partners

gene	gene	gene	gene
ABCA13	C21orf29	CAMKK1	DNAJC6
ABCC1	CACNA1C	CAMSAP1	DYRK3
ABCC12	CACNA1G	CAMTA1	EIF2C2
ABCC6	CNTNAP4	CAP2	FAM184B
ABL1	CUL3	CCDC147	FREM2
ADAM12	DMD	CCDC158	GDPD2
ADCY10	DUSP27	CELF2	GLI3
ADCY2	ECE1	CILP	IL1RN
ADCY8	EYS	CMYA5	ISX
AGBL4	FAM172A	COL14A1	KIDINS220
AHNAK	FAM184B	CORO7	LRBA
ANXA7	FGFR4	CSMD2	LY75
AP4S1	ITGAV	CUL3	MDH2
AQP2	LRP1	DDI2	MMP12
ARMC6	LY75	DEPDC5	N4BP2L2
ATP5B	MAPKAP1	DEPDC7	NCF2
ATP6AP1L	MYT1	DI10L	NCOR1
ATP6V0D2	NCF2	DMD	NCRNA00157
ATXN1	NCOR1	EDA	NRXN3
BAHD1	NHSL2	EFHC1	PARP16
BBX	NKAIN2	EFS	PLA2G2F
BCA10	NR3C1	EIF2C2	PLEK2
C15orf23	NUP188	ENTPD2	PRKCH
C15orf33	OSBPL10	EYS	PTPRS
C21orf29	PACSIN1	FAM160A1	ROBO1
C2CD3	PARP16	MUSK	SASH3
C6orf170	PDZRN4	NEUROG1	SH3BP5
C7orf44	POLM	NHSL2	SLC44A2
CACNA1C	PPP1R3A	NR3C1	SLC5A4
CACNA1G	PSEN1	ODZ1	SNX5
FAM168A	PTPRD	PCDH12	SORCS2
FAM172A	PTPRS	PLCL1	SRRM1

gene	gene	gene	gene
FAM192A	RALYL	PLEKHM3	SSX3
FAM19A2	RERE	PLOD3	STAG2
FBXL4	RIMBP2	PRKCH	STK24
FH	RNF216	PSEN1	SURF6
FREM2	SDAD1	SEPT5	SYNPO2
GAPVD1	SEC14L3	SLC44A2	TAF1
GLI3	SH3RF3	SNTA1	TMEM80
GPR182	SLC9A1	USP48	TNFRSF10B
GSTA3	SMOC2	VSNL1	TTYH1
IGFBP3	SNX5	WDFY1	UNC93B1
ITGA9	TACC2	WISP2	VSNL1
ITGB2	SRGAP1	XRRA1	XRCC4
JOSD2	SSX3	LRRC4B	ZNF410
KIDINS220	SUMF1	LRRK2	TRIOBP
LAMA2	SYNPO2	MAPKAP1	TTYH1
LCLAT1	TNFRSF10B	MST1R	LRBA
LIN9			

[00154] The nucleic acid can be any type of nucleic acid, including genomic DNA, complementary DNA (cDNA), recombinant DNA, synthetic or semi-synthetic DNA, as well as any form of corresponding RNA. A cDNA is a form of DNA artificially synthesized from a messenger RNA template and is used to produce gene clones. A synthetic DNA is free of modifications that can be found in cellular nucleic acids, including, but not limited to, histones and methylation. For example, a nucleic acid encoding a FGFR fusion molecule can comprise a recombinant nucleic acid encoding such a protein. The nucleic acid can be a non-naturally occurring nucleic acid created artificially (such as by assembling, cutting, ligating or amplifying sequences). It can be double-stranded or single-stranded.

[00155] The invention further provides for nucleic acids that are complementary to a FGFR fusion molecule. Complementary nucleic acids can hybridize to the nucleic acid sequence described above under stringent hybridization conditions. Non-limiting examples of stringent hybridization conditions include temperatures above 30°C, above 35°C, in excess of 42°C, and/or salinity of less than about 500 mM, or less than 200 mM. Hybridization conditions can be adjusted by the skilled artisan via modifying the temperature, salinity and/or the concentration of other reagents such as SDS or SSC.

[00156] According to the invention, protein variants can include amino acid sequence modifications. For example, amino acid sequence modifications fall into one or more of three classes: substitutional, insertional or deletional variants. Insertions can include amino and/or carboxyl terminal fusions as well as intrasequence insertions of single or multiple amino acid residues. Insertions ordinarily will be smaller insertions than those of amino or carboxyl terminal fusions, for example, on the order of one to four residues. Deletions are characterized by the removal of one or more amino acid residues from the protein sequence. These variants ordinarily are prepared by site-specific mutagenesis of nucleotides in the DNA encoding the protein, thereby producing DNA encoding the variant, and thereafter expressing the DNA in recombinant cell culture.

[00157] In one embodiment, a FGFR fusion molecule comprises a protein or polypeptide encoded by a nucleic acid sequence encoding a FGFR fusion molecule, such as the sequences shown in SEQ ID NOS: 80-82, 84, 94-144, or 145. In some embodiments, the nucleic acid sequence encoding a FGFR fusion molecule is about 70%, about 75%, about 80%, about 85%, about 90%, about 93%, about 95%, about 97%, about 98%, or about 99% identical to SEQ ID NOS: 80-82, 84, 94-144, or 145. In another embodiment, the polypeptide can be modified, such as by glycosylations and/or acetylations and/or chemical reaction or coupling, and can contain one or several non-natural or synthetic amino acids. An example of a FGFR fusion molecule is the polypeptide having the amino acid sequence shown in SEQ ID NOS: 79, 88, 150, 158-160, or 161. In some embodiments, the FGFR fusion molecule that is a polypeptide is about 70%, about 75%, about 80%, about 85%, about 90%, about 93%, about 95%, about 97%, about 98%, or about 99% identical to SEQ ID NOS: 79, 88, 150, 158-160, or 161. In another embodiment, a FGFR fusion molecule can be a fragment of a FGFR fusion protein. For example, the FGFR fusion molecule can encompass any portion of at least about 8 consecutive amino acids of SEQ ID NOS: 79, 88, 150, 158-160, or 161. The fragment can comprise at least about 10 amino acids, a least about 20 amino acids, at least about 30 amino acids, at least about 40 amino acids, a least about 50 amino acids, at least about 60 amino acids, or at least about 75 amino acids of SEQ ID NOS: 79, 88, 150, 158-160, or 161. Fragments include all possible amino acid lengths between about 8 and 100 about amino acids, for example, lengths between about 10 and 100 amino acids, between about 15 and 100 amino acids, between about 20 and 100 amino acids, between about 35 and 100 amino acids, between about 40 and 100 amino acids, between about 50 and 100 amino acids, between about 70 and 100 amino acids, between about 75 and 100 amino acids, or between

about 80 and 100 amino acids. Fragments include all possible amino acid lengths between about 100 and 800 amino acids, for example, lengths between about 125 and 800 amino acids, between about 150 and 800 amino acids, between about 175 and 800 amino acids, between about 200 and 800 amino acids, between about 225 and 800 amino acids, between about 250 and 800 amino acids, between about 275 and 800 amino acids, between about 300 and 800 amino acids, between about 325 and 800 amino acids, between about 350 and 800 amino acids, between about 375 and 800 amino acids, between about 400 and 800 amino acids, between about 425 and 800 amino acids, between about 450 and 800 amino acids, between about 475 and 800 amino acids, between about 500 and 800 amino acids, between about 525 and 800 amino acids, between about 550 and 800 amino acids, between about 575 and 800 amino acids, between about 600 and 800 amino acids, between about 625 and 800 amino acids, between about 650 and 800 amino acids, between about 675 and 800 amino acids, between about 700 and 800 amino acids, between about 725 and 800 amino acids, between about 750 and 800 amino acids, or between about 775 and 800 amino acids.

[00158] Chemical Synthesis. Nucleic acid sequences encoding a FGFR fusion molecule can be synthesized, in whole or in part, using chemical methods known in the art. Alternatively, a polypeptide can be produced using chemical methods to synthesize its amino acid sequence, such as by direct peptide synthesis using solid-phase techniques. Protein synthesis can either be performed using manual techniques or by automation. Automated synthesis can be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer).

[00159] Optionally, polypeptides fragments can be separately synthesized and combined using chemical methods to produce a full-length molecule. For example, these methods can be utilized to synthesize a fusion protein of the invention. In one embodiment, the fusion protein comprises a tyrosine kinase domain of an FGFR protein fused to a polypeptide that constitutively activates the tyrosine kinase domain of the FGFR protein. In another embodiment, a fusion protein comprises a transforming acidic coiled-coil (TACC) domain fused to a polypeptide with a tyrosine kinase domain, wherein the TACC domain constitutively activates the tyrosine kinase domain. An example of a fusion protein is the FGFR1-TACC1 polypeptide, which comprises the amino acid sequence shown in SEQ ID NO: 150. An example of a fusion protein is the FGFR3-TACC3 polypeptide, which has the amino acid sequence comprising SEQ ID NO: 79, 158, 159, 160, or 161.

[00160] Obtaining, Purifying and Detecting FGFR fusion molecules. A polypeptide encoded by a nucleic acid, such as a nucleic acid encoding a FGFR fusion molecule, or a variant thereof, can be obtained by purification from human cells expressing a protein or polypeptide encoded by such a nucleic acid. Non-limiting purification methods include size exclusion chromatography, ammonium sulfate fractionation, ion exchange chromatography, affinity chromatography, and preparative gel electrophoresis.

[00161] A synthetic polypeptide can be substantially purified via high performance liquid chromatography (HPLC), such as ion exchange chromatography (IEX-HPLC). The composition of a synthetic polypeptide, such as a FGFR fusion molecule, can be confirmed by amino acid analysis or sequencing.

[00162] Other constructions can also be used to join a nucleic acid sequence encoding a polypeptide/protein of the claimed invention to a nucleotide sequence encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, Wash.). Including cleavable linker sequences (i.e., those specific for Factor Xa or enterokinase (Invitrogen, San Diego, Calif.)) between the purification domain and a polypeptide encoded by a nucleic acid of the invention also can be used to facilitate purification. For example, the skilled artisan can use an expression vector encoding 6 histidine residues that precede a thioredoxin or an enterokinase cleavage site in conjunction with a nucleic acid of interest. The histidine residues facilitate purification by immobilized metal ion affinity chromatography, while the enterokinase cleavage site provides a means for purifying the polypeptide encoded by, for example, an FGFR1-TACC1, FGFR2-TACC2, FGFR3-TACC3, other FGFR-TACC, FGFR-containing, or TACC containing nucleic acid.

[00163] Host cells which contain a nucleic acid encoding a FGFR fusion molecule, and which subsequently express the same, can be identified by various procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques which include membrane, solution, or chip-based technologies for the detection and/or quantification of nucleic acid or protein. For example, the presence of a nucleic acid encoding a FGFR fusion molecule can

be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes or fragments of nucleic acids encoding the same. In one embodiment, a nucleic acid fragment of a FGFR fusion molecule can encompass any portion of at least about 8 consecutive nucleotides of SEQ ID NOS: 80-82, 84, 94-144, or 145. In another embodiment, the fragment can comprise at least about 10 consecutive nucleotides, at least about 15 consecutive nucleotides, at least about 20 consecutive nucleotides, or at least about 30 consecutive nucleotides of SEQ ID NOS: 80-82, 84, 94-144, or 145. Fragments can include all possible nucleotide lengths between about 8 and about 100 nucleotides, for example, lengths between about 15 and about 100 nucleotides, or between about 20 and about 100 nucleotides. Nucleic acid amplification-based assays involve the use of oligonucleotides selected from sequences encoding a FGFR fusion molecule nucleic acid, or FGFR fusion molecule nucleic acid to detect transformants which contain a nucleic acid encoding a protein or polypeptide of the same.

[00164] Protocols are known in the art for detecting and measuring the expression of a polypeptide encoded by a nucleic acid, such as a nucleic acid encoding a FGFR fusion molecule, using either polyclonal or monoclonal antibodies specific for the polypeptide. Non-limiting examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay using monoclonal antibodies reactive to two non-interfering epitopes on a polypeptide encoded by a nucleic acid, such as a nucleic acid encoding a FGFR fusion molecule, can be used, or a competitive binding assay can be employed.

[00165] Labeling and conjugation techniques are known by those skilled in the art and can be used in various nucleic acid and amino acid assays. Methods for producing labeled hybridization or PCR probes for detecting sequences related to nucleic acid sequences encoding a protein, such as FGFR fusion molecule, include, but are not limited to, oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, nucleic acid sequences, such as nucleic acids encoding a FGFR fusion molecule, can be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and can be used to synthesize RNA probes *in vitro* by addition of labeled nucleotides and an appropriate RNA polymerase such as T7, T3, or SP6. These procedures can be conducted using a variety of commercially available kits (Amersham Pharmacia Biotech, Promega, and US Biochemical). Suitable reporter molecules

or labels which can be used for ease of detection include radionuclides, enzymes, and fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, and/or magnetic particles.

[00166] A fragment can be a fragment of a protein, such as a FGFR fusion protein. For example, a fragment of a FGFR fusion molecule can encompass any portion of at least about 8 consecutive amino acids of SEQ ID NOS: 79, 88, 150, 158-160, or 161. The fragment can comprise at least about 10 consecutive amino acids, at least about 20 consecutive amino acids, at least about 30 consecutive amino acids, at least about 40 consecutive amino acids, a least about 50 consecutive amino acids, at least about 60 consecutive amino acids, at least about 70 consecutive amino acids, at least about 75 consecutive amino acids, at least about 80 consecutive amino acids, at least about 85 consecutive amino acids, at least about 90 consecutive amino acids, at least about 95 consecutive amino acids, at least about 100 consecutive amino acids, at least about 200 consecutive amino acids, at least about 300 consecutive amino acids, at least about 400 consecutive amino acids, at least about 500 consecutive amino acids, at least about 600 consecutive amino acids, at least about 700 consecutive amino acids, or at least about 800 consecutive amino acids of SEQ ID NOS: 79, 88, 150, 158-160, or 161. Fragments include all possible amino acid lengths between about 8 and 100 about amino acids, for example, lengths between about 10 and about 100 amino acids, between about 15 and about 100 amino acids, between about 20 and about 100 amino acids, between about 35 and about 100 amino acids, between about 40 and about 100 amino acids, between about 50 and about 100 amino acids, between about 70 and about 100 amino acids, between about 75 and about 100 amino acids, or between about 80 and about 100 amino acids.

Cell Transfection

[00167] Host cells transformed with a nucleic acid sequence of interest can be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The polypeptide produced by a transformed cell can be secreted or contained intracellularly depending on the sequence and/or the vector used. Expression vectors containing a nucleic acid sequence, such as a nucleic acid encoding a FGFR fusion molecule, can be designed to contain signal sequences which direct secretion of soluble polypeptide molecules encoded by the nucleic acid. Cell transfection and culturing methods are described in more detail below.

[00168] A eukaryotic expression vector can be used to transfect cells in order to produce proteins encoded by nucleotide sequences of the vector, e.g. those encoding a FGFR fusion molecule. Mammalian cells can contain an expression vector (for example, one that contains a nucleic acid encoding a fusion protein comprising a tyrosine kinase domain of an FGFR protein fused to a polypeptide that constitutively activates the tyrosine kinase domain of the FGFR protein, or a nucleic acid encoding a fusion protein comprises a transforming acidic coiled-coil (TACC) domain fused to a polypeptide with a tyrosine kinase domain, wherein the TACC domain constitutively activates the tyrosine kinase domain) via introducing the expression vector into an appropriate host cell via methods known in the art.

[00169] A host cell strain can be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed polypeptide encoded by a nucleic acid, in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the polypeptide also can be used to facilitate correct insertion, folding and/or function. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38), are available from the American Type Culture Collection (ATCC; 10801 University Boulevard, Manassas, Va. 20110-2209) and can be chosen to ensure the correct modification and processing of the foreign protein.

[00170] An exogenous nucleic acid can be introduced into a cell via a variety of techniques known in the art, such as lipofection, microinjection, calcium phosphate or calcium chloride precipitation, DEAE-dextran-mediated transfection, or electroporation. Electroporation is carried out at approximate voltage and capacitance to result in entry of the DNA construct(s) into cells of interest (such as glioma cells (cell line SF188), neuroblastoma cells (cell lines IMR-32, SK-N-SH, SH-F and SH-N), astrocytes and the like). Other transfection methods also include modified calcium phosphate precipitation, polybrene precipitation, liposome fusion, and receptor-mediated gene delivery.

[00171] Cells that will be genetically engineered can be primary and secondary cells obtained from various tissues, and include cell types which can be maintained and propagated in culture. Non-limiting examples of primary and secondary cells include epithelial cells, neural cells, endothelial cells, glial cells, fibroblasts, muscle cells (such as myoblasts)

keratinocytes, formed elements of the blood (e.g., lymphocytes, bone marrow cells), and precursors of these somatic cell types.

[00172] Vertebrate tissue can be obtained by methods known to one skilled in the art, such a punch biopsy or other surgical methods of obtaining a tissue source of the primary cell type of interest. In one embodiment, a punch biopsy or removal (e.g., by aspiration) can be used to obtain a source of cancer cells (for example, glioma cells, neuroblastoma cells, and the like). A mixture of primary cells can be obtained from the tissue, using methods readily practiced in the art, such as explanting or enzymatic digestion (for examples using enzymes such as pronase, trypsin, collagenase, elastase dispase, and chymotrypsin). Biopsy methods have also been described in United States Patent No. 7,419,661 and PCT application publication WO 2001/32840, and each are hereby incorporated by reference.

[00173] Primary cells can be acquired from the individual to whom the genetically engineered primary or secondary cells are administered. However, primary cells can also be obtained from a donor, other than the recipient, of the same species. The cells can also be obtained from another species (for example, rabbit, cat, mouse, rat, sheep, goat, dog, horse, cow, bird, or pig). Primary cells can also include cells from a purified vertebrate tissue source grown attached to a tissue culture substrate (for example, flask or dish) or grown in a suspension; cells present in an explant derived from tissue; both of the aforementioned cell types plated for the first time; and cell culture suspensions derived from these plated cells. Secondary cells can be plated primary cells that are removed from the culture substrate and replated, or passaged, in addition to cells from the subsequent passages. Secondary cells can be passaged one or more times. These primary or secondary cells can contain expression vectors having a gene that encodes a FGFR fusion molecule.

Cell Culturing

[00174] Various culturing parameters can be used with respect to the host cell being cultured. Appropriate culture conditions for mammalian cells are well known in the art (Cleveland WL, et al., *J Immunol Methods*, 1983, 56(2): 221-234) or can be determined by the skilled artisan (see, for example, Animal Cell Culture: A Practical Approach 2nd Ed., Rickwood, D. and Hames, B. D., eds. (Oxford University Press: New York, 1992)). Cell culturing conditions can vary according to the type of host cell selected. Commercially available medium can be utilized. Non-limiting examples of medium include, for example,

Minimal Essential Medium (MEM, Sigma, St. Louis, Mo.); Dulbecco's Modified Eagles Medium (DMEM, Sigma); Ham's F10 Medium (Sigma); HyClone cell culture medium (HyClone, Logan, Utah); RPMI-1640 Medium (Sigma); and chemically-defined (CD) media, which are formulated for various cell types, e.g., CD-CHO Medium (Invitrogen, Carlsbad, Calif.).

[00175] The cell culture media can be supplemented as necessary with supplementary components or ingredients, including optional components, in appropriate concentrations or amounts, as necessary or desired. Cell culture medium solutions provide at least one component from one or more of the following categories: (1) an energy source, usually in the form of a carbohydrate such as glucose; (2) all essential amino acids, and usually the basic set of twenty amino acids plus cysteine; (3) vitamins and/or other organic compounds required at low concentrations; (4) free fatty acids or lipids, for example linoleic acid; and (5) trace elements, where trace elements are defined as inorganic compounds or naturally occurring elements that can be required at very low concentrations, usually in the micromolar range.

[00176] The medium also can be supplemented electively with one or more components from any of the following categories: (1) salts, for example, magnesium, calcium, and phosphate; (2) hormones and other growth factors such as, serum, insulin, transferrin, and epidermal growth factor; (3) protein and tissue hydrolysates, for example peptone or peptone mixtures which can be obtained from purified gelatin, plant material, or animal byproducts; (4) nucleosides and bases such as, adenosine, thymidine, and hypoxanthine; (5) buffers, such as HEPES; (6) antibiotics, such as gentamycin or ampicillin; (7) cell protective agents, for example pluronic polyol; and (8) galactose. In one embodiment, soluble factors can be added to the culturing medium.

[00177] The mammalian cell culture that can be used with the present invention is prepared in a medium suitable for the type of cell being cultured. In one embodiment, the cell culture medium can be any one of those previously discussed (for example, MEM) that is supplemented with serum from a mammalian source (for example, fetal bovine serum (FBS)). In another embodiment, the medium can be a conditioned medium to sustain the growth of host cells.

[00178] Three-dimensional cultures can be formed from agar (such as Gey's Agar), hydrogels (such as matrigel, agarose, and the like; Lee et al., (2004) *Biomaterials* 25: 2461-

2466) or polymers that are cross-linked. These polymers can comprise natural polymers and their derivatives, synthetic polymers and their derivatives, or a combination thereof. Natural polymers can be anionic polymers, cationic polymers, amphipathic polymers, or neutral polymers. Non-limiting examples of anionic polymers can include hyaluronic acid, alginic acid (alginate), carageenan, chondroitin sulfate, dextran sulfate, and pectin. Some examples of cationic polymers, include but are not limited to, chitosan or polylysine. (Peppas et al., (2006) *Adv Mater.* 18: 1345-60; Hoffman, A. S., (2002) *Adv Drug Deliv Rev.* 43: 3-12; Hoffman, A. S., (2001) *Ann NY Acad Sci* 944: 62-73). Examples of amphipathic polymers can include, but are not limited to collagen, gelatin, fibrin, and carboxymethyl chitin. Non-limiting examples of neutral polymers can include dextran, agarose, or pullulan. (Peppas et al., (2006) *Adv Mater.* 18: 1345-60; Hoffman, A. S., (2002) *Adv Drug Deliv Rev.* 43: 3-12; Hoffman, A. S., (2001) *Ann NY Acad Sci* 944: 62-73).

[00179] Cells to be cultured can harbor introduced expression vectors, such as plasmids. The expression vector constructs can be introduced via transformation, microinjection, transfection, lipofection, electroporation, or infection. The expression vectors can contain coding sequences, or portions thereof, encoding the proteins for expression and production. Expression vectors containing sequences encoding the produced proteins and polypeptides, as well as the appropriate transcriptional and translational control elements, can be generated using methods well known to and practiced by those skilled in the art. These methods include synthetic techniques, *in vitro* recombinant DNA techniques, and *in vivo* genetic recombination which are described in J. Sambrook et al., 2001, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, N.Y. and in F. M. Ausubel et al., 1989, Current Protocols in Molecular Biology, John Wiley & Sons, New York, N.Y.

FGFR Fusion Molecule Inhibitors

[00180] The invention provides methods for use of compounds that decrease the expression level or activity of a FGFR fusion molecule in a subject. In addition, the invention provides methods for using compounds for the treatment of a gene-fusion associated cancer. In one embodiment, the gene-fusion associated cancer is an epithelial cancer. In one embodiment, the gene-fusion associated cancer comprises glioblastoma multiforme, breast cancer, lung cancer, prostate cancer, or colorectal carcinoma.

[00181] As used herein, a "FGFR fusion molecule inhibitor" refers to a compound that interacts with a FGFR fusion molecule of the invention and modulates its activity and/or its expression. For example, the compound can decrease the activity or expression of a FGFR fusion molecule. The compound can be an antagonist of a FGFR fusion molecule (e.g., a FGFR fusion molecule inhibitor). Some non-limiting examples of FGFR fusion molecule inhibitors include peptides (such as peptide fragments comprising a FGFR fusion molecule, or antibodies or fragments thereof), small molecules, and nucleic acids (such as siRNA or antisense RNA specific for a nucleic acid comprising a FGFR fusion molecule). Antagonists of a FGFR fusion molecule decrease the amount or the duration of the activity of an FGFR fusion protein. In one embodiment, the fusion protein comprises a tyrosine kinase domain of an FGFR protein fused to a polypeptide that constitutively activates the tyrosine kinase domain of the FGFR protein (e.g., FGFR1-TACC1, FGFR2-TACC2, FGFR3-TACC3 or other FGFR-TACC), or a fusion protein comprises a transforming acidic coiled-coil (TACC) domain fused to a polypeptide with a tyrosine kinase domain, wherein the TACC domain constitutively activates the tyrosine kinase domain. Antagonists include proteins, nucleic acids, antibodies, small molecules, or any other molecule which decrease the activity of a FGFR fusion molecule.

[00182] The term "modulate," as it appears herein, refers to a change in the activity or expression of a FGFR fusion molecule. For example, modulation can cause a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of a FGFR fusion molecule, such as an FGFR fusion protein.

[00183] In one embodiment, a FGFR fusion molecule inhibitor can be a peptide fragment of a FGFR fusion protein that binds to the protein itself.

[00184] For example, the FGFR fusion polypeptide can encompass any portion of at least about 8 consecutive amino acids of SEQ ID NOS: 79, 88, 150, 158-160, or 161. The fragment can comprise at least about 10 consecutive amino acids, at least about 20 consecutive amino acids, at least about 30 consecutive amino acids, at least about 40 consecutive amino acids, a least about 50 consecutive amino acids, at least about 60 consecutive amino acids, at least about 70 consecutive amino acids, at least about 75 consecutive amino acids, at least about 80 consecutive amino acids, at least about 85 consecutive amino acids, at least about 90 consecutive amino acids, at least about 95 consecutive amino acids, at least about 100 consecutive amino acids, at least about 200

consecutive amino acids, at least about 300 consecutive amino acids, at least about 400 consecutive amino acids, at least about 500 consecutive amino acids, at least about 600 consecutive amino acids, at least about 700 consecutive amino acids, or at least about 800 consecutive amino acids of SEQ ID NOS: 79, 88, 150, 158-160, or 161. Fragments include all possible amino acid lengths between about 8 and 100 about amino acids, for example, lengths between about 10 and about 100 amino acids, between about 15 and about 100 amino acids, between about 20 and about 100 amino acids, between about 35 and about 100 amino acids, between about 40 and about 100 amino acids, between about 50 and about 100 amino acids, between about 70 and about 100 amino acids, between about 75 and about 100 amino acids, or between about 80 and about 100 amino acids. These peptide fragments can be obtained commercially or synthesized via liquid phase or solid phase synthesis methods (Atherton et al., (1989) Solid Phase Peptide Synthesis: a Practical Approach. IRL Press, Oxford, England). The FGFR fusion peptide fragments can be isolated from a natural source, genetically engineered, or chemically prepared. These methods are well known in the art.

[00185] A FGFR fusion molecule inhibitor can be a protein, such as an antibody (monoclonal, polyclonal, humanized, chimeric, or fully human), or a binding fragment thereof, directed against a FGFR fusion moleculeof the invention. An antibody fragment can be a form of an antibody other than the full-length form and includes portions or components that exist within full-length antibodies, in addition to antibody fragments that have been engineered. Antibody fragments can include, but are not limited to, single chain Fv (scFv), diabodies, Fv, and $(Fab')_2$, triabodies, Fc, Fab, CDR1, CDR2, CDR3, combinations of CDR's, variable regions, tetrabodies, bifunctional hybrid antibodies, framework regions, constant regions, and the like (*see*, Maynard et al., (2000) *Ann. Rev. Biomed. Eng.* 2:339-76; Hudson (1998) *Curr. Opin. Biotechnol.* 9:395-402). Antibodies can be obtained commercially, custom generated, or synthesized against an antigen of interest according to methods established in the art (*see* United States Patent Nos. 6,914,128, 5,780,597, 5,811,523; Roland E. Kontermann and Stefan Dübel (editors), Antibody Engineering, Vol. I & II, (2010) 2nd ed., Springer; Antony S. Dimitrov (editor), Therapeutic Antibodies: Methods and Protocols (Methods in Molecular Biology), (2009), Humana Press; Benny Lo (editor) Antibody Engineering: Methods and Protocols (Methods in Molecular Biology), (2004) Humana Press, each of which are hereby incorporated by reference in their entireties). For example, antibodies directed to a FGFR fusion molecule can be obtained commercially from Abcam, Santa Cruz Biotechnology, Abgent, R&D Systems, Novus Biologicals, etc. Human

antibodies directed to a FGFR fusion molecule (such as monoclonal, humanized, fully human, or chimeric antibodies) can be useful antibody therapeutics for use in humans. In one embodiment, an antibody or binding fragment thereof is directed against SEQ ID NOS: 79, 88, 150, 158-160, or 161.

[00186] Inhibition of RNA encoding a FGFR fusion molecule can effectively modulate the expression of a FGFR fusion molecule. Inhibitors are selected from the group comprising: siRNA; interfering RNA or RNAi; dsRNA; RNA Polymerase III transcribed DNAs; ribozymes; and antisense nucleic acids, which can be RNA, DNA, or an artificial nucleic acid.

[00187] Antisense oligonucleotides, including antisense DNA, RNA, and DNA/RNA molecules, act to directly block the translation of mRNA by binding to targeted mRNA and preventing protein translation. For example, antisense oligonucleotides of at least about 15 bases and complementary to unique regions of the DNA sequence encoding a FGFR fusion molecule can be synthesized, e.g., by conventional phosphodiester techniques (Dallas et al., (2006) *Med. Sci. Monit.*12(4):RA67-74; Kalota et al., (2006) *Handb. Exp. Pharmacol.* 173:173-96; Lutzelburger et al., (2006) *Handb. Exp. Pharmacol.* 173:243-59). Antisense nucleotide sequences include, but are not limited to: morpholinos, 2'-O-methyl polynucleotides, DNA, RNA and the like.

[00188] siRNA comprises a double stranded structure containing from about 15 to about 50 base pairs, for example from about 21 to about 25 base pairs, and having a nucleotide sequence identical or nearly identical to an expressed target gene or RNA within the cell. The siRNA comprise a sense RNA strand and a complementary antisense RNA strand annealed together by standard Watson-Crick base-pairing interactions. The sense strand comprises a nucleic acid sequence which is substantially identical to a nucleic acid sequence contained within the target miRNA molecule. "Substantially identical" to a target sequence contained within the target mRNA refers to a nucleic acid sequence that differs from the target sequence by about 3% or less. The sense and antisense strands of the siRNA can comprise two complementary, single-stranded RNA molecules, or can comprise a single molecule in which two complementary portions are base-paired and are covalently linked by a single-stranded "hairpin" area. See also, McMnaus and Sharp (2002) *Nat Rev Genetics*, 3:737-47, and Sen and Blau (2006) *FASEB J.*, 20:1293-99, the entire disclosures of which are herein incorporated by reference.

[00189] The siRNA can be altered RNA that differs from naturally-occurring RNA by the addition, deletion, substitution and/or alteration of one or more nucleotides. Such alterations can include addition of non-nucleotide material, such as to the end(s) of the siRNA or to one or more internal nucleotides of the siRNA, or modifications that make the siRNA resistant to nuclease digestion, or the substitution of one or more nucleotides in the siRNA with deoxyribo-nucleotides. One or both strands of the siRNA can also comprise a 3' overhang. As used herein, a 3' overhang refers to at least one unpaired nucleotide extending from the 3'-end of a duplexed RNA strand. For example, the siRNA can comprise at least one 3' overhang of from 1 to about 6 nucleotides (which includes ribonucleotides or deoxyribonucleotides) in length, or from 1 to about 5 nucleotides in length, or from 1 to about 4 nucleotides in length, or from about 2 to about 4 nucleotides in length. For example, each strand of the siRNA can comprise 3' overhangs of dithymidylic acid ("TT") or diuridylic acid ("uu").

[00190] siRNA can be produced chemically or biologically, or can be expressed from a recombinant plasmid or viral vector (for example, see U.S. Patent No. 7,294,504 and U.S. Patent No. 7,422,896, the entire disclosures of which are herein incorporated by reference). Exemplary methods for producing and testing dsRNA or siRNA molecules are described in U.S. Patent Application Publication No. 2002/0173478 to Gewirtz, U.S. Patent No. 8,071,559 to Hannon et al., and in U.S. Patent No. 7,148,342 to Tolentino et al., the entire disclosures of which are herein incorporated by reference.

[00191] In one embodiment, an siRNA directed to a human nucleic acid sequence comprising a FGFR fusion molecule can be generated against any one of SEQ ID NOS: 80-82, 84, 94-144, or 145. In another embodiment, an siRNA directed to a human nucleic acid sequence comprising a breakpoint of anFGFR fusion molecule can be generated against any one of SEQ ID NOS: 1-77, 80-82, 84-144, or 145. In one embodiment, the hairpin sequences targeting the *FGFR3* gene comprise SEQ ID NOS: 183, 184, or 185.

[00192] RNA polymerase III transcribed DNAs contain promoters, such as the U6 promoter. These DNAs can be transcribed to produce small hairpin RNAs in the cell that can function as siRNA or linear RNAs, which can function as antisense RNA. The FGFR fusion molecule inhibitor can comprise ribonucleotides, deoxyribonucleotides, synthetic nucleotides, or any suitable combination such that the target RNA and/or gene is inhibited. In addition, these forms of nucleic acid can be single, double, triple, or quadruple stranded. (see for

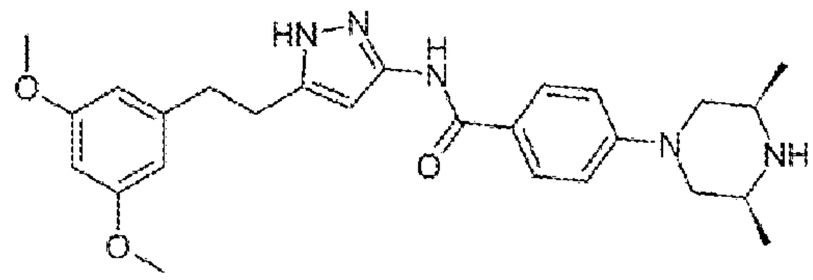
example Bass (2001) *Nature*, 411:428-429; Elbashir et al., (2001) *Nature*, 411:494 498; U.S. Patent No. 6,509,154; U.S. Patent Application Publication No. 2003/0027783; and PCT Publication Nos. WO 00/044895, WO 99/032619, WO 00/01846, WO 01/029058, WO 00/044914).

[00193] FGFR fusion molecule inhibitor can be a small molecule that binds to a FGFR fusion protein described herein and disrupts its function. Small molecules are a diverse group of synthetic and natural substances generally having low molecular weights. They can be isolated from natural sources (for example, plants, fungi, microbes and the like), are obtained commercially and/or available as libraries or collections, or synthesized. Candidate small molecules that inhibit a FGFR fusion protein can be identified via *in silico* screening or high-through-put (HTP) screening of combinatorial libraries according to methods established in the art (e.g., see Potyrailo et al., (2011) *ACS Comb Sci.* 13(6):579-633; Mensch et al., (2009) *J Pharm Sci.* 98(12):4429-68; Schnur (2008) *Curr Opin Drug Discov Devel.* 11(3):375-80; and Jhoti (2007) *Ernst Schering Found Symp Proc.* (3):169-85, each of which are hereby incorporated by reference in their entireties.) Most conventional pharmaceuticals, such as aspirin, penicillin, and many chemotherapeutics, are small molecules, can be obtained commercially, can be chemically synthesized, or can be obtained from random or combinatorial libraries as described below (*see*, e.g., Werner et al., (2006) *Brief Funct. Genomic Proteomic* 5(1):32-6).

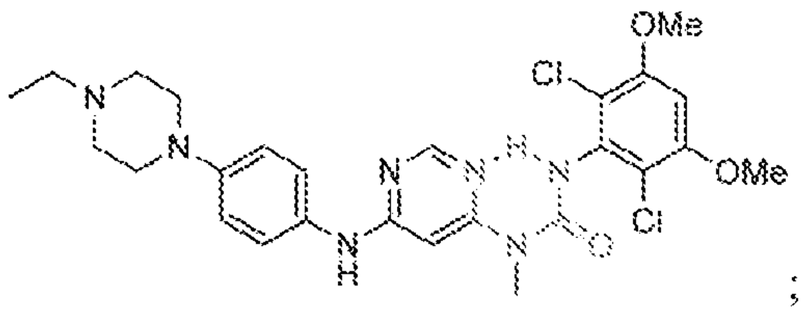
[00194] Non-limiting examples of FGFR fusion molecule inhibitors include the FGFR inhibitors AZD4547 (*see* Gavine et al., (2012) *Cancer Res*, 72(8); 2045–56; *see also* PCT Application Publication No. WO 2008/075068, each of which are hereby incorporated by reference in their entireties); NVP-BGJ398 (*see* Guagnano et al., (2011) *J. Med. Chem.*, 54:7066–7083; *see also* U.S. Patent Application Publication No. 2008-0312248 A1, each of which are hereby incorporated by reference in their entireties); PD173074 (*see* Guagnano et al., (2011) *J. Med. Chem.*, 54:7066–7083; see also Mohammadi et al., (1998) *EMBO J.*, 17:5896–5904, each of which are hereby incorporated by reference in their entireties); NF449 (EMD Millipore (Billerica, MA) Cat. No. 480420; *see also* Krejci, (2010) *the Journal of Biological Chemistry*, 285(27):20644–20653, which is hereby incorporated by reference in its entirety); LY2874455 (Active Biochem; *see* Zhao et al. (2011) *Mol Cancer Ther.* (11):2200-10; *see also* PCT Application Publication No. WO 2010129509, each of which are hereby incorporated by reference in their entireties); TKI258 (Dovitinib); BIBF-1120 (Intedanib-

Vargatef); BMS-582664 (Brivanib alaninate); AZD-2171 (Cediranib); TSU-68 (Orantinib); AB-1010 (Masitinib); AP-24534 (Ponatinib); and E-7080 (by Eisai). A non-limiting example of an FGFR fusion molecule inhibitor includes the TACC inhibitor KHS101 (Wurdak et al., (2010) *PNAS*, 107(38): 16542-47, which is hereby incorporated by reference in its entirety).

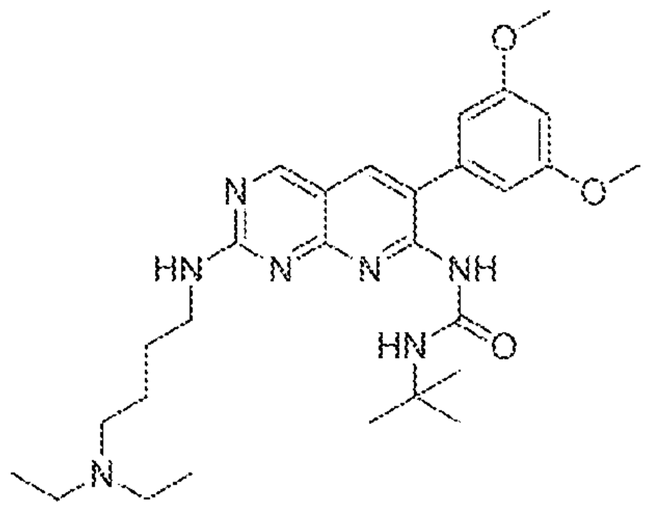
[00195] Structures of FGFR fusion molecule inhibitors useful for the invention include, but are not limited to: the FGFR inhibitor AZD4547,



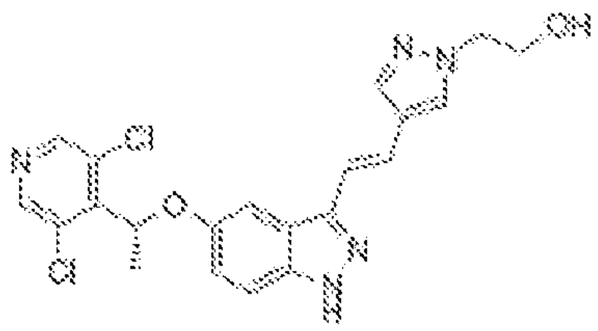
; the FGFR inhibitor NVP-BGJ398,



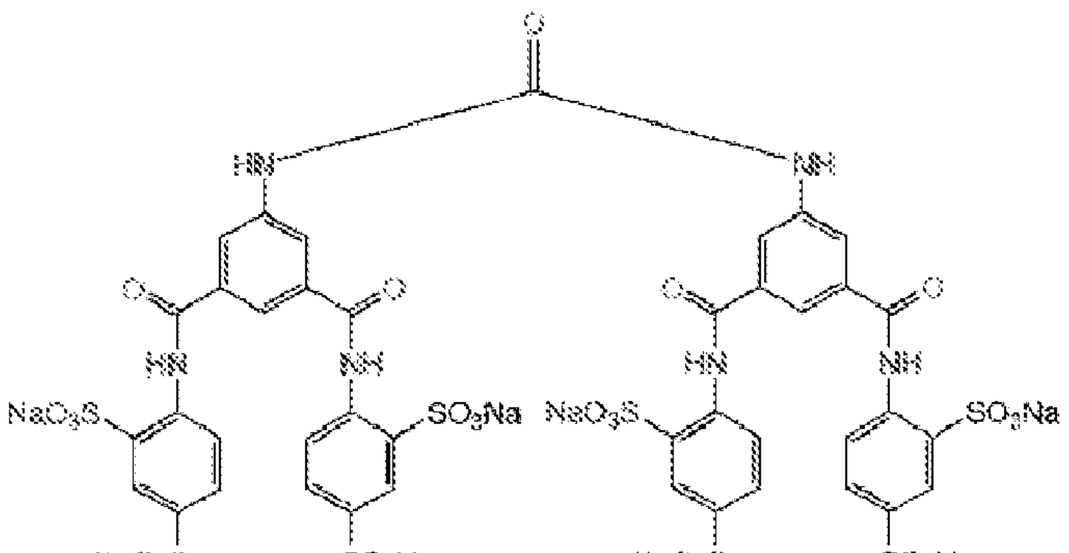
; the FGFR inhibitor PD173074,



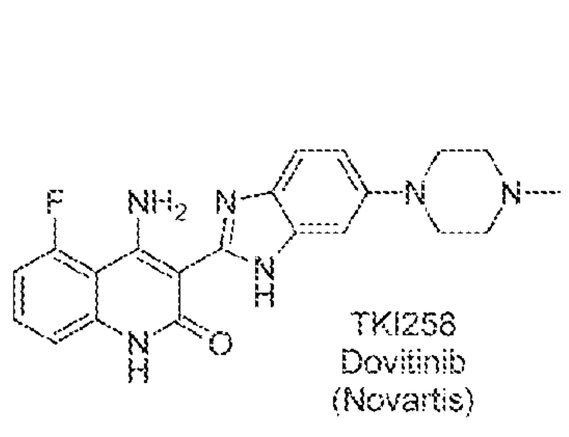
; the FGFR inhibitor LY2874455



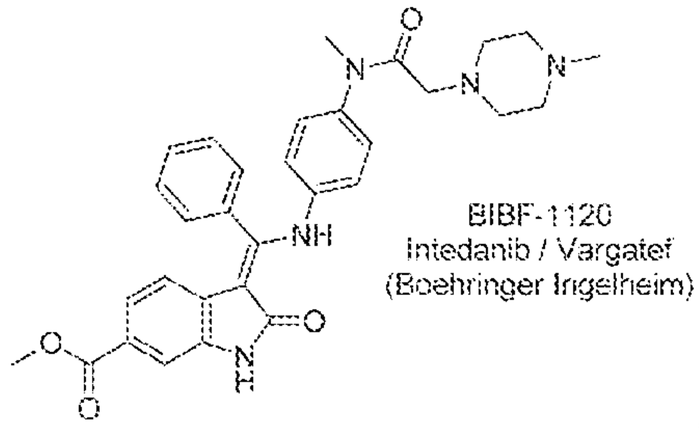
; and the FGFR inhibitor NF449 (EMD Millipore (Billerica, MA) Cat. No. 480420),



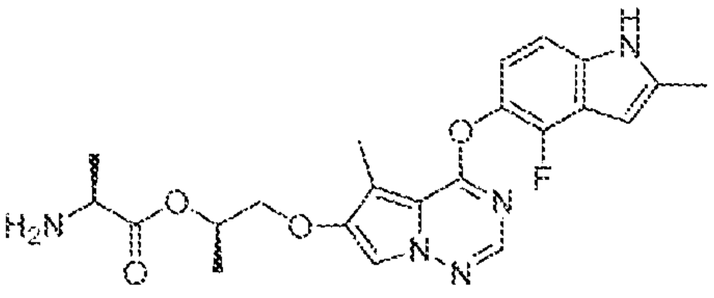
[00196] Other FGFR inhibitors include, but are not limited to:



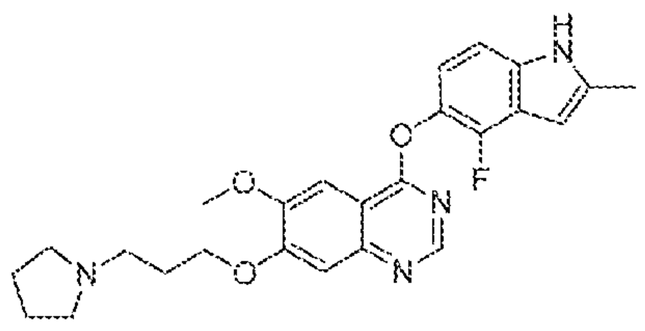
TKI258
Dovitinib
(Novartis)



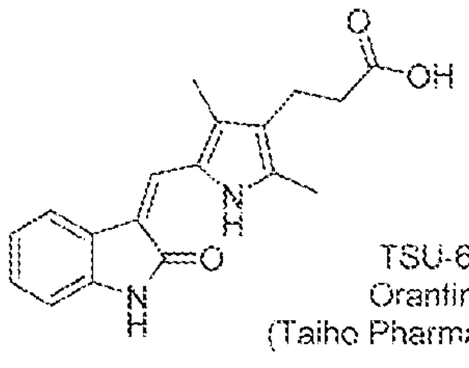
BIBF-1120
Intedanib / Vargatef
(Boehringer Ingelheim)



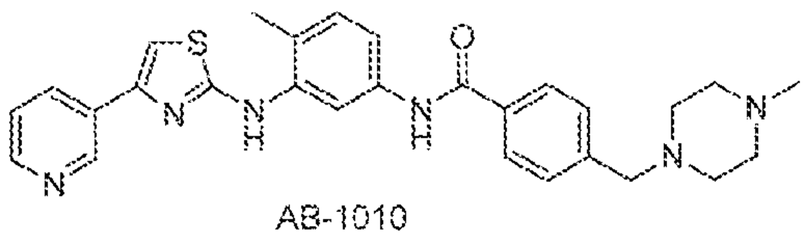
BMS-582664
Brivanib alaninate
(Bristol-Myers Squibb)



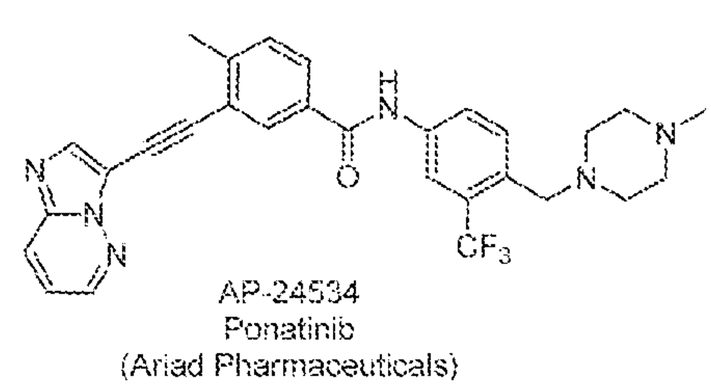
AZD-2171
Cediranib
(AstraZeneca)



TSU-68
Orantinib
(Taiho Pharmaceutical)

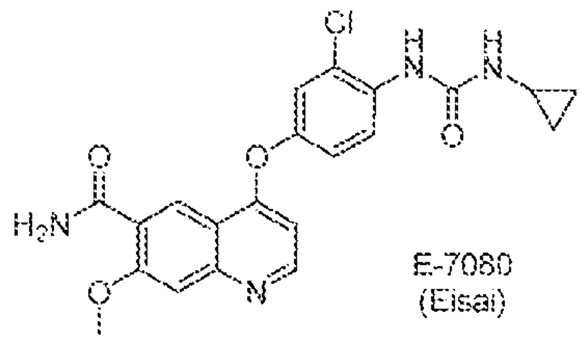


AB-1010
Masitinib
(AB Science)



AP-24534
Ponatinib
(Ariad Pharmaceuticals)

, and

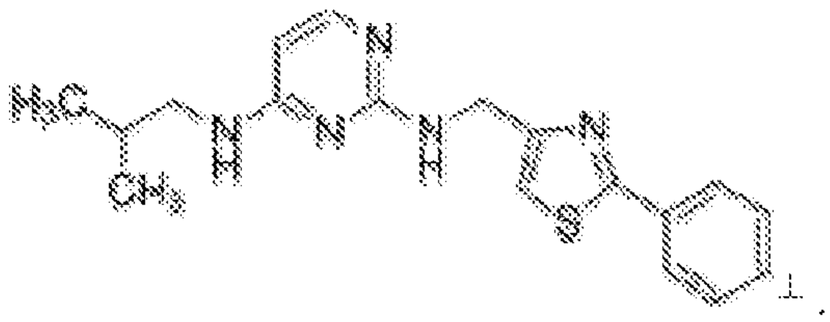


E-7080
(Eisai)

.

[00197] A structure of an FGFR fusion molecule inhibitor useful for the invention include,

but is not limited to the TACC inhibitor KHS101,



Assessment and Therapuetic Treatment

[00198] The invention provides a method of decreasing the growth of a solid tumor in a subject. The tumor is associated with, but not limited to, glioblastoma multiforme, breast cancer, lung cancer, prostate cancer, or colorectal carcinoma. In one embodiment, the method comprises detecting the presence of a FGFR fusion molecule in a sample obtained from a subject. In some embodiments, the sample is incubated with an agent that binds to an FGFR fusion molecule, such as an antibody, a probe, a nucleic acid primer, and the like. In further embodiments, the method comprises administering to the subject an effective amount of a FGFR fusion molecule inhibitor, wherein the inhibitor decreases the size of the solid tumor.

[00199] The invention also provides a method for treating or preventing a gene-fusion associated cancer in a subject. In one embodiment, the gene-fusion associated cancer comprises an epithelial cancer. In one embodiment, the gene-fusion associated cancer comprises glioblastoma multiforme, breast cancer, lung cancer, prostate cancer, or colorectal carcinoma. In some embodiments, the epithelial cancer comprises bladder urothelial carcinoma, breast carcinoma, colorectal cancer, prostate carcinoma, lung squamous cell carcinoma, head and neck squamous cell carcinoma, or a combination of the epithelial cancers decribed. In one embodiment, the method comprises detecting the presence of a FGFR fusion molecule in a sample obtained from a subject, the presence of the fusion being indicative of a gene-fusion associated cancer, and, administering to the subject in need a therapeutic treatment against a gene-fusion associated cancer. In some embodiments, the sample is incubated with an agent that binds to an FGFR fusion molecule, such as an antibody, a probe, a nucleic acid primer, and the like.

[00200] The invention also provides a method for decreasing in a subject in need thereof the expression level or activity of a fusion protein comprising the tyrosine kinase domain of

an FGFR protein fused to a polypeptide that constitutively activates the tyrosine kinase domain of the FGFR protein. In some embodiments, the method comprises obtaining a biological sample from the subject. In some embodiments, the sample is incubated with an agent that binds to an FGFR fusion molecule, such as an antibody, a probe, a nucleic acid primer, and the like. In some embodiments, the method comprises administering to the subject a therapeutic amount of a composition comprising an admixture of a pharmaceutically acceptable carrier an inhibitor of the fusion protein of the invention. In another embodiment, the method further comprises determining the fusion protein expression level or activity. In another embodiment, the method further comprises detecting whether the fusion protein expression level or activity is decreased as compared to the fusion protein expression level or activity prior to administration of the composition, thereby decreasing the expression level or activity of the fusion protein. In some embodiments, the fusion protein is an FGFR-TACC fusion protein.

[00201] The administering step in each of the claimed methods can comprise a drug administration, such as FGFR fusion molecule inhibitor (for example, a pharmaceutical composition comprising an antibody that specifically binds to a FGFR fusion molecule or a fragment thereof; a small molecule that specifically binds to a FGFR protein; a small molecule that specifically binds to a TACC protein; an antisense RNA or antisense DNA that decreases expression of a FGFR fusion molecule; a siRNA that specifically targets a gene encoding a FGFR fusion molecule; or a combination thereof). In one embodiment, the therapeutic molecule to be administered comprises a polypeptide of a FGFR fusion molecule, comprising at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 93%, at least about 95%, at least about 97%, at least about 98%, at least about 99%, or 100% of the amino acid sequence of SEQ ID NOS: 79, 88, 150, 158-160, or 161, and exhibits the function of decreasing expression of such a protein, thus treating a gene fusion-associated cancer. In another embodiment, administration of the therapeutic molecule decreases the size of the solid tumor associated with glioblastoma multiforme, breast cancer, lung cancer, prostate cancer, or colorectal carcinoma.

[00202] In another embodiment, the therapeutic molecule to be administered comprises an siRNA directed to a human nucleic acid sequence comprising a FGFR fusion molecule. In one embodiment, the siRNA is directed to any one of SEQ ID NOS: 80-82, 84, 94-144, or 145. In another embodiment, the siRNA is directed to any one of SEQ ID NOS: 1-77, 80-82,

84-144, or 145. In a further embodiment, the therapeutic molecule to be administered comprises an antibody or binding fragment thereof, which is directed against SEQ ID NOS: 79, 88, 150, 158-160, or 161. In some embodiments, the therapeutic molecule to be administered comprises a small molecule that specifically binds to a FGFR protein, such as AZD4547, NVP-BGJ398, PD173074, NF449, TK1258, BIBF-1120, BMS-582664, AZD-2171, TSU68, AB1010, AP24534, E-7080, or LY2874455. In other embodiments, the therapeutic molecule to be administered comprises a small molecule that specifically binds to a TACC protein, such as KHS101.

[00203] In one embodiment, the invention provides for the detection of a chromosomal rearrangement at given chromosomal coordinates. In another embodiment, the detection or determination comprises nucleic acid sequencing, selective hybridization, selective amplification, gene expression analysis, or a combination thereof. In another embodiment, the detection or determination comprises protein expression analysis, for example by western blot analysis, ELISA, or other antibody detection methods.

[00204] In one embodiment, the biological sample comprises neuronal cells, serum, bone marrow, blood, peripheral blood, lymph nodes, cerebro-spinal fluid, urine, a saliva sample, a buccal swab, a serum sample, a sputum sample, a lacrimal secretion sample, a semen sample, a vaginal secretion sample, a fetal tissue sample, or a combination thereof.

[00205] A FGFR fusion molecule, for example, a fusion between FGFR1, FGFR2, FGFR3, or any other FGFR, and TACC1, TACC2, TACC3 or any other TACC, can be determined at the level of the DNA, RNA, or polypeptide. Optionally, detection can be determined by performing an oligonucleotide ligation assay, a confirmation based assay, a hybridization assay, a sequencing assay, an allele-specific amplification assay, a microsequencing assay, a melting curve analysis, a denaturing high performance liquid chromatography (DHPLC) assay (for example, *see* Jones et al, (2000) *Hum Genet.*, 106(6):663-8), or a combination thereof. In one embodiment, the detection is performed by sequencing all or part of a FGFR fusion molecule (e.g., a FGFR1-TACC1, FGFR2-TACC2, FGFR3-TACC3 or other FGFR-TACC nucleic acid, or a FGFR1, TACC1, FGFR2, TACC2, FGFR3, TACC3 or other FGFR or TACC nucleic acid), or by selective hybridization or amplification of all or part of a FGFR fusion molecule (e.g., a FGFR1-TACC1, FGFR2-TACC2, FGFR3-TACC3 or other FGFR-TACC nucleic acid, or a FGFR1, TACC1, FGFR2, TACC2, FGFR3, TACC3 or other FGFR or TACC nucleic acid). A FGFR fusion molecule

specific amplification (e.g., a FGFR1-TACC1, FGFR2-TACC2, FGFR3-TACC3 or other FGFR-TACC nucleic acid specific amplification) can be carried out before the fusion identification step.

[00206] The invention provides for a method of detecting a chromosomal alteration in a subject afflicted with a gene-fusion associated cancer. In one embodiment, the chromosomal alteration is an in-frame fused transcript described herein, for example an FGFR fusion molecule. In some embodiments, the chromosomal alteration is a chromosomal translocation, for example an FGFR fusion molecule. An alteration in a chromosome region occupied by a FGFR fusion molecule, such as a FGFR1-TACC1, FGFR2-TACC2, FGFR3-TACC3 or other FGFR-TACC nucleic acid, can be any form of mutation(s), deletion(s), rearrangement(s) and/or insertions in the coding and/or non-coding region of the locus, alone or in various combination(s). Mutations can include point mutations. Insertions can encompass the addition of one or several residues in a coding or non-coding portion of the gene locus. Insertions can comprise an addition of between 1 and 50 base pairs in the gene locus. Deletions can encompass any region of one, two or more residues in a coding or non-coding portion of the gene locus, such as from two residues up to the entire gene or locus. Deletions can affect smaller regions, such as domains (introns) or repeated sequences or fragments of less than about 50 consecutive base pairs, although larger deletions can occur as well. Rearrangement includes inversion of sequences. The alteration in a chromosome region occupied by a FGFR fusion molecule, e.g., a FGFR1-TACC1, FGFR2-TACC2, FGFR3-TACC3 or other FGFR-TACC nucleic acid, can result in amino acid substitutions, RNA splicing or processing, product instability, the creation of stop codons, production of oncogenic fusion proteins, frame-shift mutations, and/or truncated polypeptide production. The alteration can result in the production of a FGFR fusion molecule, for example, one encoded by a FGFR1-TACC1, FGFR2-TACC2, FGFR3-TACC3 or other FGFR-TACC nucleic acid, with altered function, stability, targeting or structure. The alteration can also cause a reduction, or even an increase in protein expression. In one embodiment, the alteration in the chromosome region occupied by a FGFR fusion molecule can comprise a chromosomal rearrangement resulting in the production of a FGFR fusion molecule, such as a FGFR1-TACC1, FGFR2-TACC2, FGFR3-TACC3 or other FGFR-TACC fusion. This alteration can be determined at the level of the DNA, RNA, or polypeptide. In another embodiment, the detection or determination comprises nucleic acid sequencing, selective hybridization, selective amplification, gene expression analysis, or a combination thereof. In

another embodiment, the detection or determination comprises protein expression analysis, for example by western blot analysis, ELISA, or other antibody detection methods. In one embodiment, the coordinates comprising FGFR1 translocations comprise chr8:38,268,656-38,325,363. In another embodiment, the coordinates comprising FGFR2 translocations comprise chr10:123,237,844-123,357,972. In a further embodiment, the coordinates comprising FGFR3 translocations comprise chr4:1,795,039-1,810,599. In yet another embodiment, the coordinates comprising FGFR4 translocations comprise chr5:176,513,921-176,525,126. In one embodiment, the coordinates comprising TACC1 translocations comprise chr8:38,644,722-38,710,546. In another embodiment, the coordinates comprising TACC2 translocations comprise chr10:123,748,689-124,014,057. In a further embodiment, the coordinates comprising TACC3 translocations comprise chr4:1,723,217-1,746,905.

[00207] The present invention provides a method for treating a gene-fusion associated cancer in a subject in need thereof. In one embodiment, the method comprises obtaining a sample from the subject to determine the level of expression of an FGFR fusion molecule in the subject. In some embodiments, the sample is incubated with an agent that binds to an FGFR fusion molecule, such as an antibody, a probe, a nucleic acid primer, and the like. In another embodiment, the detection or determination comprises nucleic acid sequencing, selective hybridization, selective amplification, gene expression analysis, or a combination thereof. In another embodiment, the detection or determination comprises protein expression analysis, for example by western blot analysis, ELISA, or other antibody detection methods. In some embodiments, the method further comprises assessing whether to administer a FGFR fusion molecule inhibitor based on the expression pattern of the subject. In further embodiments, the method comprises administering a FGFR fusion molecule inhibitor to the subject. In one embodiment, the gene-fusion associated cancer comprises an epithelial cancer. In one embodiment, the gene-fusion associated cancer comprises glioblastoma multiforme, breast cancer, lung cancer, prostate cancer, or colorectal carcinoma. In some embodiments, the epithelial cancer comprises bladder urothelial carcinoma, breast carcinoma, colorectal cancer, prostate carcinoma, lung squamous cell carcinoma, head and neck squamous cell carcinoma, or a combination of the epithelial cancers decribed.

[00208] In one embodiment, the invention provides for a method of detecting the presence of altered RNA expression of an FGFR fusion molecule in a subject, for example one afflicted with a gene-fusion associated cancer. In another embodiment, the invention

provides for a method of detecting the presence of an FGFR fusion molecule in a subject. In some embodiments, the method comprises obtaining a sample from the subject to determine whether the subject expresses an FGFR fusion molecule. In some embodiments, the sample is incubated with an agent that binds to an FGFR fusion molecule, such as an antibody, a probe, a nucleic acid primer, and the like. In other embodiments, the detection or determination comprises nucleic acid sequencing, selective hybridization, selective amplification, gene expression analysis, or a combination thereof. In another embodiment, the detection or determination comprises protein expression analysis, for example by western blot analysis, ELISA, or other antibody detection methods. In some embodiments, the method further comprises assessing whether to administer a FGFR fusion molecule inhibitor based on the expression pattern of the subject. In further embodiments, the method comprises administering a FGFR fusion molecule inhibitor to the subject. Altered RNA expression includes the presence of an altered RNA sequence, the presence of an altered RNA splicing or processing, or the presence of an altered quantity of RNA. These can be detected by various techniques known in the art, including sequencing all or part of the RNA or by selective hybridization or selective amplification of all or part of the RNA. In a further embodiment, the method can comprise detecting the presence or expression of a FGFR fusion molecule, such as one encoded by a FGFR1-TACC1, FGFR2-TACC2, FGFR3-TACC3 or other FGFR-TACC nucleic acid. Altered polypeptide expression includes the presence of an altered polypeptide sequence, the presence of an altered quantity of polypeptide, or the presence of an altered tissue distribution. These can be detected by various techniques known in the art, including by sequencing and/or binding to specific ligands (such as antibodies). In one embodiment, the detecting comprises using a northern blot; real time PCR and primers directed to SEQ ID NOS: 80-82, 84, 94-144, or 145; a ribonuclease protection assay; a hybridization, amplification, or sequencing technique to detect an FGFR fusion molecule, such as one comprising SEQ ID NOS: 80-82, 84, 94-144, or 145; or a combination thereof. In another embodiment, the PCR primers comprise SEQ ID NOS: 162, 163, 164, or 165. In a further embodiment, primers used for the screening of FGFR fusion molecules, such as FGFR-TACC fusions, comprise SEQ ID NOS: 166, 167, 168, or 169. In some embodiments, primers used for genomic detection of an FGFR3-TACC3 fusion comprise SEQ ID NOS: 170 and 171.

[00209] Various techniques known in the art can be used to detect or quantify altered gene or RNA expression or nucleic acid sequences, which include, but are not limited to,

hybridization, sequencing, amplification, and/or binding to specific ligands (such as antibodies). Other suitable methods include allele-specific oligonucleotide (ASO), oligonucleotide ligation, allele-specific amplification, Southern blot (for DNAs), Northern blot (for RNAs), single-stranded conformation analysis (SSCA), PFGE, fluorescent in situ hybridization (FISH), gel migration, clamped denaturing gel electrophoresis, denaturing HLPC, melting curve analysis, heteroduplex analysis, RNase protection, chemical or enzymatic mismatch cleavage, ELISA, radio-immunoassays (RIA) and immuno-enzymatic assays (IEMA).

[00210] Some of these approaches (such as SSCA and constant gradient gel electrophoresis (CGGE)) are based on a change in electrophoretic mobility of the nucleic acids, as a result of the presence of an altered sequence. According to these techniques, the altered sequence is visualized by a shift in mobility on gels. The fragments can then be sequenced to confirm the alteration. Some other approaches are based on specific hybridization between nucleic acids from the subject and a probe specific for wild type or altered gene or RNA. The probe can be in suspension or immobilized on a substrate. The probe can be labeled to facilitate detection of hybrids. Some of these approaches are suited for assessing a polypeptide sequence or expression level, such as Northern blot, ELISA and RIA. These latter require the use of a ligand specific for the polypeptide, for example, the use of a specific antibody.

[00211] Hybridization. Hybridization detection methods are based on the formation of specific hybrids between complementary nucleic acid sequences that serve to detect nucleic acid sequence alteration(s). A detection technique involves the use of a nucleic acid probe specific for a wild type or altered gene or RNA, followed by the detection of the presence of a hybrid. The probe can be in suspension or immobilized on a substrate or support (for example, as in nucleic acid array or chips technologies). The probe can be labeled to facilitate detection of hybrids. In one embodiment, the probe according to the invention can comprise a nucleic acid directed to SEQ ID NOS: 80-82, 84, 94-144, or 145. For example, a sample from the subject can be contacted with a nucleic acid probe specific for a gene encoding a FGFR fusion molecule, and the formation of a hybrid can be subsequently assessed. In one embodiment, the method comprises contacting simultaneously the sample with a set of probes that are specific for an FGFR fusion molecule. Also, various samples from various subjects can be investigated in parallel.

[00212] According to the invention, a probe can be a polynucleotide sequence which is complementary to and specifically hybridizes with a, or a target portion of a, gene or RNA corresponding to a FGFR fusion molecule. Useful probes are those that are complementary to the gene, RNA, or target portion thereof. Probes can comprise single-stranded nucleic acids of between 8 to 1000 nucleotides in length, for instance between 10 and 800, between 15 and 700, or between 20 and 500. Longer probes can be used as well. A useful probe of the invention is a single stranded nucleic acid molecule of between 8 to 500 nucleotides in length, which can specifically hybridize to a region of a gene or RNA that corresponds to a FGFR fusion molecule.

[00213] The sequence of the probes can be derived from the sequences of the FGFR fusion genes provided herein. Nucleotide substitutions can be performed, as well as chemical modifications of the probe. Such chemical modifications can be accomplished to increase the stability of hybrids (e.g., intercalating groups) or to label the probe. Some examples of labels include, without limitation, radioactivity, fluorescence, luminescence, and enzymatic labeling.

[00214] A guide to the hybridization of nucleic acids is found in e.g., Sambrook, ed., Molecular Cloning: A Laboratory Manual (3rd Ed.), Vols. 1-3, Cold Spring Harbor Laboratory, 1989; Current Protocols In Molecular Biology, Ausubel, ed. John Wiley & Sons, Inc., New York, 2001; Laboratory Techniques In Biochemistry And Molecular Biology: Hybridization With Nucleic Acid Probes, Part I. Theory and Nucleic Acid Preparation, Tijssen, ed. Elsevier, N.Y., 1993.

[00215] Sequencing. Sequencing can be carried out using techniques well known in the art, using automatic sequencers. The sequencing can be performed on the complete FGFR fusion molecule or on specific domains thereof.

[00216] Amplification. Amplification is based on the formation of specific hybrids between complementary nucleic acid sequences that serve to initiate nucleic acid reproduction. Amplification can be performed according to various techniques known in the art, such as by polymerase chain reaction (PCR), ligase chain reaction (LCR), strand displacement amplification (SDA) and nucleic acid sequence based amplification (NASBA). These techniques can be performed using commercially available reagents and protocols. Useful techniques in the art encompass real-time PCR, allele-specific PCR, or PCR based

single-strand conformational polymorphism (SSCP). Amplification usually requires the use of specific nucleic acid primers, to initiate the reaction. For example, nucleic acid primers useful for amplifying sequences corresponding to a FGFR fusion molecule are able to specifically hybridize with a portion of the gene locus that flanks a target region of the locus. In one embodiment, amplification comprises using forward and reverse PCR primers directed to SEQ ID NOS: 80-82, 84, 94-144, or 145. Nucleic acid primers useful for amplifying sequences from a FGFR fusion molecule (e.g., a FGFR1-TACC1, FGFR2-TACC2, FGFR3-TACC3 or other FGFR-TACC nucleic acid); the primers specifically hybridize with a portion of an FGFR fusion molecule. In certain subjects, the presence of an FGFR fusion molecule corresponds to a subject with a gene fusion-associated cancer. In one embodiment, amplification can comprise using forward and reverse PCR primers comprising nucleotide sequences of SEQ ID NOS: 80-82, 84, 94-144, or 145.

[00217] Non-limiting amplification methods include, e.g., polymerase chain reaction, PCR (PCR Protocols, A Guide To Methods And Applications, ed. Innis, Academic Press, N.Y., 1990 and PCR Strategies, 1995, ed. Innis, Academic Press, Inc., N.Y.); ligase chain reaction (LCR) (Wu (1989) *Genomics* 4:560; Landegren (1988) *Science* 241:1077; Barringer (1990) *Gene* 89:117); transcription amplification (Kwoh (1989) *PNAS* 86:1173); and, self-sustained sequence replication (Guatelli (1990) *PNAS* 87:1874); Q Beta replicase amplification (Smith (1997) *J. Clin. Microbiol.* 35:1477-1491), automated Q-beta replicase amplification assay (Burg (1996) *Mol. Cell. Probes* 10:257-271) and other RNA polymerase mediated techniques (e.g., NASBA, Cangene, Mississauga, Ontario; *see also* Berger (1987) *Methods Enzymol.* 152:307-316; U.S. Pat. Nos. 4,683,195 and 4,683,202; and Sooknanan (1995) *Biotechnology* 13:563-564). All the references stated above are incorporated by reference in their entireties.

[00218] The invention provides for a nucleic acid primer, wherein the primer can be complementary to and hybridize specifically to a portion of a FGFR fusion molecule, such as a FGFR1-TACC1, FGFR2-TACC2, FGFR3-TACC3 or other FGFR-TACC nucleic acid (e.g., DNA or RNA) in certain subjects having a gene fusion-associated cancer. In one embodiment, the gene-fusion associated cancer comprises glioblastoma multiforme, breast cancer, lung cancer, prostate cancer, or colorectal carcinoma. Primers of the invention can be specific for fusion sequences in a FGFR1-TACC1, FGFR2-TACC2, FGFR3-TACC3 or other FGFR-TACC nucleic acid (DNA or RNA). By using such primers, the detection of an amplification product indicates the presence of a fusion of a FGFR1 and TACC1, FGFR2 and

TACC2, FGFR3 and TACC3 or other FGFR and TACC nucleic acid. Examples of primers of this invention can be single-stranded nucleic acid molecules of about 5 to 60 nucleotides in length, or about 8 to about 25 nucleotides in length. The sequence can be derived directly from the sequence of a FGFR fusion molecule, e.g. FGFR1-TACC1, FGFR2-TACC2, FGFR3-TACC3 or other FGFR-TACC nucleic acid. Perfect complementarity is useful to ensure high specificity; however, certain mismatch can be tolerated. For example, a nucleic acid primer or a pair of nucleic acid primers as described above can be used in a method for detecting the presence of a gene fusion-associated cancer in a subject. In one embodiment, primers can be used to detect an FGFR fusion molecule, such as a primer comprising SEQ ID NOS: 80-82, 84, 94-144, or 145; or a combination thereof. In another embodiment, the PCR primers comprise SEQ ID NOS: 162, 163, 164, or 165. In a further embodiment, primers used for the screening of FGFR fusion molecules, such as FGFR-TACC fusions, comprise SEQ ID NOS: 166, 167, 168, or 169. In some embodiments, primers used for genomic detection of an FGFR3-TACC3 fusion comprise SEQ ID NOS: 170 and 171.

[00219] Specific Ligand Binding. As discussed herein, a nucleic acid encoding a FGFR fusion molecule or expression of a FGFR fusion molecule, can also be detected by screening for alteration(s) in a sequence or expression level of a polypeptide encoded by the same. Different types of ligands can be used, such as specific antibodies. In one embodiment, the sample is contacted with an antibody specific for a polypeptide encoded by a FGFR fusion molecule and the formation of an immune complex is subsequently determined. Various methods for detecting an immune complex can be used, such as ELISA, radioimmunoassays (RIA) and immuno-enzymatic assays (IEMA).

[00220] For example, an antibody can be a polyclonal antibody, a monoclonal antibody, as well as fragments or derivatives thereof having substantially the same antigen specificity. Fragments include Fab, Fab'2, or CDR regions. Derivatives include single-chain antibodies, humanized antibodies, or poly-functional antibodies. An antibody specific for a polypeptide encoded by a FGFR fusion molecule can be an antibody that selectively binds such a polypeptide. In one embodiment, the antibody is raised against a polypeptide encoded by a FGFR fusion molecule (such as FGFR1-TACC1, FGFR2-TACC2, FGFR3-TACC3 or other FGFR-TACC fusion) or an epitope-containing fragment thereof. Although non-specific binding towards other antigens can occur, binding to the target polypeptide occurs with a higher affinity and can be reliably discriminated from non-specific binding. In one

embodiment, the method can comprise contacting a sample from the subject with an antibody specific for a FGFR fusion molecule, and determining the presence of an immune complex. Optionally, the sample can be contacted to a support coated with antibody specific for a FGFR fusion molecule. In one embodiment, the sample can be contacted simultaneously, or in parallel, or sequentially, with various antibodies specific for different forms of a FGFR fusion molecule, e.g., FGFR1-TACC1, FGFR2-TACC2, FGFR3-TACC3 or other FGFR-TACC fusion.

[00221] The invention also provides for a diagnostic kit comprising products and reagents for detecting in a sample from a subject the presence of a FGFR fusion molecule. The kit can be useful for determining whether a sample from a subject exhibits increased or reduced expression of a FGFR fusion molecule. For example, the diagnostic kit according to the present invention comprises any primer, any pair of primers, any nucleic acid probe and/or any ligand, or any antibody directed specifically to a FGFR fusion molecule. The diagnostic kit according to the present invention can further comprise reagents and/or protocols for performing a hybridization, amplification, or antigen-antibody immune reaction. In one embodiment, the kit can comprise nucleic acid primers that specifically hybridize to and can prime a polymerase reaction from a FGFR fusion molecule comprising SEQ ID NOS: 80-82, 84, 94-144, or 145, or a combination thereof. In one embodiment, primers can be used to detect a FGFR fusion molecule, such as a primer directed to SEQ ID NOS: 80-82, 84, 94-144, or 145; or a combination thereof. In another embodiment, the PCR primer comprises SEQ ID NOS: 162, 163, 164, 165, 166, 167, 168, or 169. In a further embodiment, primers used for the screening of FGFR fusion molecules, such as FGFR-TACC fusions, comprise SEQ ID NOS: 166, 167, 168, or 169. In some embodiments, primers used for genomic detection of an FGFR3-TACC3 fusion comprise SEQ ID NOS: 170 and 171. In some embodiments, the kit comprises an antibody that specifically binds to a FGFR fusion molecule comprising SEQ ID NOS: 79, 85-89, 150, 158-160, or 161, wherein the antibody will recognize the protein only when a FGFR fusion molecule is present

[00222] The diagnosis methods can be performed *in vitro*, *ex vivo*, or *in vivo*. These methods utilize a sample from the subject in order to assess the status of a FGFR fusion molecule. The sample can be any biological sample derived from a subject, which contains nucleic acids or polypeptides. Examples of such samples include, but are not limited to, fluids, tissues, cell samples, organs, and tissue biopsies. Non-limiting examples of samples

include blood, liver, plasma, serum, saliva, urine, or seminal fluid. The sample can be collected according to conventional techniques and used directly for diagnosis or stored. The sample can be treated prior to performing the method, in order to render or improve availability of nucleic acids or polypeptides for testing. Treatments include, for instance, lysis (e.g., mechanical, physical, or chemical), centrifugation. The nucleic acids and/or polypeptides can be pre-purified or enriched by conventional techniques, and/or reduced in complexity. Nucleic acids and polypeptides can also be treated with enzymes or other chemical or physical treatments to produce fragments thereof. In one embodiment, the sample is contacted with reagents, such as probes, primers, or ligands, in order to assess the presence of a FGFR fusion molecule. Contacting can be performed in any suitable device, such as a plate, tube, well, or glass. In some embodiments, the contacting is performed on a substrate coated with the reagent, such as a nucleic acid array or a specific ligand array. The substrate can be a solid or semi-solid substrate such as any support comprising glass, plastic, nylon, paper, metal, or polymers. The substrate can be of various forms and sizes, such as a slide, a membrane, a bead, a column, or a gel. The contacting can be made under any condition suitable for a complex to be formed between the reagent and the nucleic acids or polypeptides of the sample.

Nucleic Acid Delivery Methods

[00223] Delivery of nucleic acids into viable cells can be effected *ex vivo*, *in situ*, or *in vivo* by use of vectors, such as viral vectors (e.g., lentivirus, adenovirus, adeno-associated virus, or a retrovirus), or ex vivo by use of physical DNA transfer methods (e.g., liposomes or chemical treatments). Non-limiting techniques suitable for the transfer of nucleic acid into mammalian cells *in vitro* include the use of liposomes, electroporation, microinjection, cell fusion, DEAE-dextran, and the calcium phosphate precipitation method (See, for example, Anderson, *Nature*, 1998) supplement to 392(6679):25(). Introduction of a nucleic acid or a gene encoding a polypeptide of the invention can also be accomplished with extrachromosomal substrates (transient expression) or artificial chromosomes (stable expression). Cells can also be cultured ex vivo in the presence of therapeutic compositions of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced in vivo for therapeutic purposes.

[00224] Nucleic acids can be inserted into vectors and used as gene therapy vectors. A number of viruses have been used as gene transfer vectors, including papovaviruses, e.g.,

SV40 (Madzak et al., (1992) *J Gen Virol.* 73(Pt 6):1533-6), adenovirus (Berkner (1992) *Curr Top Microbiol Immunol.*158:39-66; Berkner (1988) *Biotechniques*, 6(7):616-29; Gorziglia and Kapikian (1992) *J Virol.* 66(7):4407-12; Quantin et al., (1992) *Proc Natl Acad Sci U S A.* 89(7):2581-4; Rosenfeld et al., (1992) *Cell.* 68(1):143-55; Wilkinson et al., (1992) *Nucleic Acids Res.* 20(9):2233-9; Stratford-Perricaudet et al., (1990) *Hum Gene Ther.* 1(3):241-56), vaccinia virus (Moss (1992) *Curr Opin Biotechnol.* 3(5):518-22), adeno-associated virus (Muzyczka, (1992) *Curr Top Microbiol Immunol.* 158:97-129; Ohi et al., (1990) *Gene.* 89(2):279-82), herpesviruses including HSV and EBV (Margolskee (1992) *Curr Top Microbiol Immunol.* 158:67-95; Johnson et al., (1992) *Brain Res Mol Brain Res.*12(1-3):95-102; Fink et al., (1992) *Hum Gene Ther.* 3(1):11-9; Breakefield and Geller (1987) *Mol Neurobiol.* 1(4):339-71; Freese et al., (1990) *Biochem Pharmacol.* 40(10):2189-99), and retroviruses of avian (Bandyopadhyay and Temin (1984) *Mol Cell Biol.* 4(4):749-54; Petropoulos et al., (1992) *J Virol.* 66(6):3391-7), murine (Miller et al. (1992) *Mol Cell Biol.* 12(7):3262-72; Miller et al., (1985) *J Virol.* 55(3):521-6; Sorge et al., (1984) *Mol Cell Biol.* 4(9):1730-7; Mann and Baltimore (1985) *J Virol.* 54(2):401-7; Miller et al., (1988) *J Virol.* 62(11):4337-45), and human origin (Shimada et al., (1991) *J Clin Invest.* 88(3):1043-7; Helseth et al., (1990) *J Virol.* 64(12):6314-8; Page et al., (1990) *J Virol.* 64(11):5270-6; Buchschacher and Panganiban (1992) *J Virol.* 66(5):2731-9).

[00225] Non-limiting examples of *in vivo* gene transfer techniques include transfection with viral (e.g., retroviral) vectors (*see* U.S. Pat. No. 5,252,479, which is incorporated by reference in its entirety) and viral coat protein-liposome mediated transfection (Dzau et al., (1993) *Trends in Biotechnology* 11:205-210), incorporated entirely by reference). For example, naked DNA vaccines are generally known in the art; *see* Brower, (1998) *Nature Biotechnology*, 16:1304-1305, which is incorporated by reference in its entirety. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (*see*, e.g., U.S. Pat. No. 5,328,470) or by stereotactic injection (*see*, e.g., Chen, et al., (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g., retroviral vectors, the pharmaceutical preparation can include one or more cells that produce the gene delivery system.

[00226] For reviews of nucleic acid delivery protocols and methods *see* Anderson et al. (1992) *Science* 256:808-813; U.S. Pat. Nos. 5,252,479, 5,747,469, 6,017,524, 6,143,290, 6,410,010 6,511,847; and U.S. Application Publication No. 2002/0077313, which are all hereby incorporated by reference in their entireties. For additional reviews, *see* Friedmann (1989) *Science*, 244:1275-1281; Verma, *Scientific American*: 68-84 (1990); Miller (1992) *Nature*, 357: 455-460; Kikuchi et al. (2008) *J Dermatol Sci.* 50(2):87-98; Isaka et al. (2007) *Expert Opin Drug Deliv.* 4(5):561-71; Jager et al.(2007) *Curr Gene Ther.* 7(4):272-83; Waehler et al.(2007) *Nat Rev Genet.* 8(8):573-87; Jensen et al. (2007) *Ann Med.* 39(2):108-15; Herweijer et al. (2007) *Gene Ther.* 14(2):99-107; Eliyahu et al. (2005) *Molecules* 10(1):34-64; and Altaras et al. (2005) *Adv Biochem Eng Biotechnol.* 99:193-260, all of which are hereby incorporated by reference in their entireties.

[00227] A FGFR fusion nucleic acid can also be delivered in a controlled release system. For example, the FGFR fusion molecule can be administered using intravenous infusion, an implantable osmotic pump, a transdermal patch, liposomes, or other modes of administration. In one embodiment, a pump can be used (*see* Sefton (1987) *Biomed. Eng.* 14:201; Buchwald et al. (1980) *Surgery* 88:507; Saudek et al. (1989) *N. Engl. J. Med.* 321:574). In another embodiment, polymeric materials can be used (*see* Medical Applications of Controlled Release, Langer and Wise (eds.), CRC Pres., Boca Raton, Fla. (1974); Controlled Drug Bioavailability, Drug Product Design and Performance, Smolen and Ball (eds.), Wiley, New York (1984); Ranger and Peppas, (1983) *J. Macromol. Sci. Rev. Macromol. Chem.* 23:61; *see also* Levy et al. (1985) *Science* 228:190; During et al. (1989) *Ann. Neurol.* 25:351; Howard et al. (1989) *J. Neurosurg.* 71:105). In yet another embodiment, a controlled release system can be placed in proximity of the therapeutic target thus requiring only a fraction of the systemic dose (*see*, e.g., Goodson, in Medical Applications of Controlled Release, supra, vol. 2, pp. 115-138 (1984)). Other controlled release systems are discussed in the review by Langer (*Science* (1990) 249:1527-1533).

Pharmaceutical Compositions and Administration for Therapy

[00228] An inhibitor of the invention can be incorporated into pharmaceutical compositions suitable for administration, for example the inhibitor and a pharmaceutically acceptable carrier

[00229] A FGFR fusion molecule or inhibitor of the invention can be administered to the subject once (e.g., as a single injection or deposition). Alternatively, a FGFR fusion molecule or inhibitor can be administered once or twice daily to a subject in need thereof for a period of from about two to about twenty-eight days, or from about seven to about ten days. A FGFR fusion molecule or inhibitor can also be administered once or twice daily to a subject for a period of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 times per year, or a combination thereof. Furthermore, a FGFR fusion molecule or inhibitor of the invention can be co-administrated with another therapeutic. Where a dosage regimen comprises multiple administrations, the effective amount of the FGFR fusion molecule or inhibitor administered to the subject can comprise the total amount of gene product administered over the entire dosage regimen.

[00230] A FGFR fusion molecule or inhibitor can be administered to a subject by any means suitable for delivering the FGFR fusion molecule or inhibitor to cells of the subject, such as cancer cells, e.g., glioblastoma multiforme, breast cancer, lung cancer, prostate cancer, or colorectal carcinoma. For example, a FGFR fusion molecule or inhibitor can be administered by methods suitable to transfect cells. Transfection methods for eukaryotic cells are well known in the art, and include direct injection of the nucleic acid into the nucleus or pronucleus of a cell; electroporation; liposome transfer or transfer mediated by lipophilic materials; receptor mediated nucleic acid delivery, bioballistic or particle acceleration; calcium phosphate precipitation, and transfection mediated by viral vectors.

[00231] The compositions of this invention can be formulated and administered to reduce the symptoms associated with a gene fusion-associated cancer, e.g., glioblastoma multiforme, breast cancer, lung cancer, prostate cancer, or colorectal carcinoma, by any means that produces contact of the active ingredient with the agent's site of action in the body of a subject, such as a human or animal (e.g., a dog, cat, or horse). They can be administered by any conventional means available for use in conjunction with pharmaceuticals, either as individual therapeutic active ingredients or in a combination of therapeutic active ingredients. They can be administered alone, but are generally administered with a pharmaceutical carrier selected on the basis of the chosen route of administration and standard pharmaceutical practice.

[00232] A therapeutically effective dose of FGFR fusion molecule or inhibitor can depend upon a number of factors known to those or ordinary skill in the art. The dose(s) of the

FGFR fusion molecule inhibitor can vary, for example, depending upon the identity, size, and condition of the subject or sample being treated, further depending upon the route by which the composition is to be administered, if applicable, and the effect which the practitioner desires the a FGFR fusion molecule inhibitor to have upon the nucleic acid or polypeptide of the invention. These amounts can be readily determined by a skilled artisan. Any of the therapeutic applications described herein can be applied to any subject in need of such therapy, including, for example, a mammal such as a dog, a cat, a cow, a horse, a rabbit, a monkey, a pig, a sheep, a goat, or a human.

[00233] Pharmaceutical compositions for use in accordance with the invention can be formulated in conventional manner using one or more physiologically acceptable carriers or excipients. The therapeutic compositions of the invention can be formulated for a variety of routes of administration, including systemic and topical or localized administration. Techniques and formulations generally can be found in Remmington's Pharmaceutical Sciences, Meade Publishing Co., Easton, Pa (20^{th} Ed., 2000), the entire disclosure of which is herein incorporated by reference. For systemic administration, an injection is useful, including intramuscular, intravenous, intraperitoneal, and subcutaneous. For injection, the therapeutic compositions of the invention can be formulated in liquid solutions, for example in physiologically compatible buffers such as Hank's solution or Ringer's solution. In addition, the therapeutic compositions can be formulated in solid form and redissolved or suspended immediately prior to use. Lyophilized forms are also included. Pharmaceutical compositions of the present invention are characterized as being at least sterile and pyrogen-free. These pharmaceutical formulations include formulations for human and veterinary use.

[00234] According to the invention, a pharmaceutically acceptable carrier can comprise any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Any conventional media or agent that is compatible with the active compound can be used. Supplementary active compounds can also be incorporated into the compositions.

[00235] A pharmaceutical composition containing FGFR fusion molecule inhibitor can be administered in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed herein. Such pharmaceutical compositions can comprise, for example antibodies directed to a FGFR fusion molecule, or a variant thereof, or antagonists

of a FGFR fusion molecule. The compositions can be administered alone or in combination with at least one other agent, such as a stabilizing compound, which can be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions can be administered to a patient alone, or in combination with other agents, drugs or hormones.

[00236] Sterile injectable solutions can be prepared by incorporating the FGFR fusion molecule inhibitor (e.g., a polypeptide or antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated herein, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated herein. In the case of sterile powders for the preparation of sterile injectable solutions, examples of useful preparation methods are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

[00237] In some embodiments, the FGFR fusion molecule inhibitor can be applied via transdermal delivery systems, which slowly releases the active compound for percutaneous absorption. Permeation enhancers can be used to facilitate transdermal penetration of the active factors in the conditioned media. Transdermal patches are described in for example, U.S. Pat. No. 5,407,713; U.S. Pat. No. 5,352,456; U.S. Pat. No. 5,332,213; U.S. Pat. No. 5,336,168; U.S. Pat. No. 5,290,561; U.S. Pat. No. 5,254,346; U.S. Pat. No. 5,164,189; U.S. Pat. No. 5,163,899; U.S. Pat. No. 5,088,977; U.S. Pat. No. 5,087,240; U.S. Pat. No. 5,008,110; and U.S. Pat. No. 4,921,475.

[00238] "Subcutaneous" administration can refer to administration just beneath the skin (i.e., beneath the dermis). Generally, the subcutaneous tissue is a layer of fat and connective tissue that houses larger blood vessels and nerves. The size of this layer varies throughout the body and from person to person. The interface between the subcutaneous and muscle layers can be encompassed by subcutaneous administration. This mode of administration can be feasible where the subcutaneous layer is sufficiently thin so that the factors present in the compositions can migrate or diffuse from the locus of administration. Thus, where intradermal administration is utilized, the bolus of composition administered is localized proximate to the subcutaneous layer.

[00239] Administration of the cell aggregates (such as DP or DS aggregates) is not restricted to a single route, but can encompass administration by multiple routes. For instance, exemplary administrations by multiple routes include, among others, a combination of intradermal and intramuscular administration, or intradermal and subcutaneous administration. Multiple administrations can be sequential or concurrent. Other modes of application by multiple routes will be apparent to the skilled artisan.

[00240] In other embodiments, this implantation method will be a one-time treatment for some subjects. In further embodiments of the invention, multiple cell therapy implantations will be required. In some embodiments, the cells used for implantation will generally be subject-specific genetically engineered cells. In another embodiment, cells obtained from a different species or another individual of the same species can be used. Thus, using such cells can require administering an immunosuppressant to prevent rejection of the implanted cells. Such methods have also been described in United States Patent No. 7,419,661and PCT application publication WO 2001/32840, and are hereby incorporated by reference.

[00241] A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation or ingestion), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

[00242] Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent

that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, a pharmaceutically acceptable polyol like glycerol, propylene glycol, liquid polyetheylene glycol, and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it can be useful to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

[00243] Sterile injectable solutions can be prepared by incorporating the inhibitor (e.g., a polypeptide or antibody or small molecule) of the invention in the required amount in an appropriate solvent with one or a combination of ingredients enumerated herein, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated herein. In the case of sterile powders for the preparation of sterile injectable solutions, examples of useful preparation methods are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

[00244] Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier and subsequently swallowed.

[00245] Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as

magnesium stearate or sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

[00246] Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

[00247] In some embodiments, the effective amount of the administered FGFR fusion molecule inhibitor is at least about 0.0001 μg/kg body weight, at least about 0.00025 μg/kg body weight, at least about 0.0005 μg/kg body weight, at least about 0.00075 μg/kg body weight, at least about 0.001 μg/kg body weight, at least about 0.0025 μg/kg body weight, at least about 0.005 μg/kg body weight, at least about 0.0075 μg/kg body weight, at least about 0.01 μg/kg body weight, at least about 0.025 μg/kg body weight, at least about 0.05 μg/kg body weight, at least about 0.075 μg/kg body weight, at least about 0.1 μg/kg body weight, at least about 0.25 μg/kg body weight, at least about 0.5 μg/kg body weight, at least about 0.75 μg/kg body weight, at least about 1 μg/kg body weight, at least about 5 μg/kg body weight, at least about 10 μg/kg body weight, at least about 25 μg/kg body weight, at least about 50 μg/kg body weight, at least about 75 μg/kg body weight, at least about 100 μg/kg body weight, at least about 150 μg/kg body weight, at least about 200 μg/kg body weight, at least about 250 μg/kg body weight, at least about 300 μg/kg body weight, at least about 350 μg/kg body weight, at least about 400 μg/kg body weight, at least about 450 μg/kg body weight, at least about 500 μg/kg body weight, at least about 550 μg/kg body weight, at least about 600 μg/kg body weight, at least about 650 μg/kg body weight, at least about 700 μg/kg body weight, at least about 750 μg/kg body weight, at least about 800 μg/kg body weight, at least about 850 μg/kg body weight, at least about 900 μg/kg body weight, at least about 950 μg/kg body weight, at least about 1,000 μg/kg body weight, at least about 2,000 μg/kg body weight, at least about 3,000 μg/kg body weight, at least about 4,000 μg/kg body weight, at least about 5,000 μg/kg body weight, at least about 6,000 μg/kg body weight, at least about 7,000 μg/kg

body weight, at least about 8,000 µg/kg body weight, at least about 9,500 µg/kg body weight, or at least about 10,000 µg/kg body weight.

***

[00248] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Exemplary methods and materials are described below, although methods and materials similar or equivalent to those described herein can also be used in the practice or testing of the present invention.

[00249] All publications and other references mentioned herein are incorporated by reference in their entirety, as if each individual publication or reference were specifically and individually indicated to be incorporated by reference. Publications and references cited herein are not admitted to be prior art.

EXAMPLES

[00250] Examples are provided below to facilitate a more complete understanding of the invention. The following examples illustrate the exemplary modes of making and practicing the invention. However, the scope of the invention is not limited to specific embodiments disclosed in these Examples, which are for purposes of illustration only, since alternative methods can be utilized to obtain similar results.

[00251] **Example 1**: *Transforming and recurrent fusions of FGFR and TACC gene in glioblastoma*

[00252] The history of successful targeted therapy of cancer largely coincides with the inactivation of recurrent, oncogenic and addicting gene fusions in hematological malignancies and recently in some types of epithelial cancer. Glioblastoma multiforme (GBM) is among the most lethal forms of human cancer. Here, an integrated gene fusion discovery pipeline was developed for the detection of in-frame fused transcripts from RNA-seq and genomic fusions from whole exome sequences. The application of the pipeline to human GBM unraveled recurrent chromosomal translocations, which fuse in-frame the tyrosine kinase domain of FGFR genes (FGFR1 or FGFR3) to the TACC domain of TACC1 or TACC3, respectively. The frequency of FGFR-TACC fusions is 3 of 97 GBM (3.1%). The FGFR-TACC fusion protein displays strong oncogenic activity when introduced into astrocytes or transduced by lentivirus-mediated stereotactic delivery to the adult mouse brain. The FGFR-TACC fusion protein mis-localizes over the mitotic spindle pole, has constitutive tyrosine kinase activity and dysregulates the mitotic cycle with delayed mitotic progression. The impaired mitotic fidelity triggers chromatid cohesion defects, defective spindle checkpoint activation, chromosomal mis-segregation, and rampant aneuploidy. Inhibition of FGFR kinase corrects the aneuploidy and oral administration of a specific FGFR tyrosine kinase inhibitor under clinical investigation arrests tumor growth and prolongs survival of mice harboring intracranial FGFR3-TACC3-initiated glioma. FGFR-TACC fusions identify a subset of GBM patients who may benefit from targeted inhibition of the tyrosine kinase activity of FGFR.

[00253] Glioblastoma multiforme (GBM) is among the most difficult forms of cancer to treat in humans (Ohgaki and Kleihues, 2005). So far, the targeted therapeutic approaches that have been tested against potentially important oncogenic drivers in GBM have met limited

success (Lo, 2010; Reardon et al., 2010; van den Bent et al., 2009). Recurrent chromosomal translocations leading to production of oncogenic fusion proteins are viewed as initiating and addicting events in the pathogenesis of human cancer, thus providing the most desirable molecular targets for cancer therapy (Ablain et al., 2011; Mitelman et al., 2007). Chromosomal rearrangements resulting in recurrent and oncogenic gene fusions are hallmarks of hematological malignancies and recently they have also been uncovered in subsets of solid tumors (breast, prostate, lung and colorectal carcinoma), but they have not been found in GBM (Bass et al., 2011; Prensner and Chinnaiyan, 2009). Important and successful targeted therapeutic interventions for patients whose tumors carry these rearrangements have stemmed from the discovery of functional gene fusions, especially when the translocations involve kinase-coding genes (*BCR-ABL*, *EML4-ALK*) (Druker, 2009; Gerber and Minna, 2010).

[00254] A hallmark of GBM is rampant chromosomal instability (CIN), which leads to aneuploidy (Furnari et al., 2007). CIN and aneuploidy are early events in the pathogenesis of cancer (Cahill et al., 1999). It has been suggested that genetic alterations targeting mitotic fidelity might be responsible for mis-segregation of chromosomes during mitosis, resulting in aneuploidy (Gordon et al., 2012; Solomon et al., 2011). Here, the first cases of recurrent and oncogenic gene fusions in human GBM are described. The resulting fusion protein localizes to mitotic cells, disrupts the normal control of chromosome segregation and induces aneuploidy. A therapeutic strategy with FGFR tyrosine kinase inhibitors is also reported for the targeted therapy of GBM patients harboring these chromosomal rearrangements.

[00255] ***Identification of recurrent fusions of FGFR and TACC genes.*** To identify genomic rearrangements in GBM that generate functional fusion proteins and are recurrent, gene pairs discovered as in-frame fused transcripts from the analysis of massively parallel, paired-end sequencing of expressed transcripts (RNA-seq) would also emerge as fused gene pairs from the genomic analysis of human GBM. Towards this aim, two complementary gene fusion discovery methods were devised and were applied to two GBM cohorts. The first, TX-Fuse, is an algorithm for the discovery of candidate fusion transcripts from RNA-seq (**Figure 8**). The second, Exome-Fuse, detects fusion genes from whole exome DNA sequences (**Figure 8**). As first step for the detection of fused transcripts, RNA-seq data was generated from short-term cultures of glioma stem-like cells (GSCs) freshly isolated from nine patients carrying primary GBM. The culture of primary GBM tumors under serum-free

conditions selects cells that retain phenotypes and genotypes closely mirroring primary tumor profiles as compared to serum-cultured glioma cell lines that have largely lost their developmental identities (Lee et al., 2006). Therefore, without being bound by theory, if glioma cells carry gene fusions causally responsible for the most aggressive hallmarks of GBM, they should be selected in GSCs. RNA-seq generated an average of 60.3 million paired reads for each GSC culture, of which over 80% were mapped to the reference transcriptome and genome. TX-Fuse detects two main sources of evidence: split reads and split inserts (see Experimental Procedures). The application of TX-Fuse to the RNA-seq dataset from nine GSCs led to the discovery of five candidate rearrangements (all of which were intrachromosomal) that give rise to in-frame fusion transcripts (Table 1).

[00256] **Table 1: Predicted in-frame fusion proteins from RNA-Seq of nine GSCs**

# Split Inserts	# Split Reads	Sample	Gene1	Gene2	Ref Seq1	Ref Seq2	Tx Pos1	Tx Pos2
294	76	GSC-1123	FGFR3	TACC3	NM_000142	NM_006342	2530	1751
37	54	GSC-0114	POLR2A	WRAP53	NM_000937	NM_001143990	479	798
7	48	GSC-0114	CAPZB	UBR4	NM_001206540	NM_020765	228	12111
8	29	GSC-0517	ST8SIA4	PAM	NM_005668	NM_000919	1125	730
6	17	GSC-0308	PIGU	NCOA6	NM_080476	NM_014071	729	6471
1	6	GSC-0127	IFNAR2	IL10RB	NM_000874	NM_000628	1083	149

#Split Inserts	#Split Reads	Sample	Chr 1	Strand 1	hg19 GenPos1	Chr 2	Strand 2	hg19 GenPos2
294	76	GSC-1123	4	+	1808842	4	+	1737004
37	54	GSC-0114	17	+	7399259	17	+	7604059
7	48	GSC-0114	1	-	19712098	1	-	19433440
8	29	GSC-0517	5	-	100147809	5	+	102260661
6	17	GSC-0308	20	-	33203914	20	-	33303130
1	6	GSC-0127	21	+	34632901	21	+	34640699

[00257] Next, genomic rearrangements leading to gene fusions were identified in GBM by applying Exome-Fuse to a dataset of paired-end exome DNA sequences from 84 GBM samples from TCGA (Table 2).

[00258] This analysis detected 147 paired gene fusions, thus producing an average of 1.75 gene fusion events per tumor (Table 3).

[00259] The FGFR and TACC families of genes were markedly enriched among those recurrently involved in genomic fusions, with eight tumors harboring FGFR rearrangements and seven tumors harboring fusions that implicate TACC genes (**Figure 1A**). The comparative analysis of the TX-Fuse and Exon-Fuse outputs revealed that FGFR3-TACC3 was the only fusion pair identified as either an in-frame transcript by TX-Fuse and genomic fusions by Exome-Fuse (Tables 1, 2 and 3).

[00260] **Table 2** shows fusion breakpoint information of recurrent gene fusions identified by Exome-fuse analysis of 84 GBM from TCGA. As multiple junctions may exist in each fusion candidate, information for all breakpoints is displayed. Column definitions include: sample = TCGA sample ID, virtForSplitReads/virtRevSplitReads/virtTotSplitReads = # forward/reverse/total split reads, splitInserts = # split inserts, dirA/dirB = forward (1) or reverse (0) direction of split read portion mapping to gene A/B, dirAB_matepair = direction of mate pair of split read, cosmicA+B = # recorded mutations of gene A+B in COSMIC.

Table 2Fusion breakpoint information of recurrent gene fusions identified by Exome-fuse analysis of 84 GBM from TCGA.

Sample	virtForSplitReads	virtRevSplitReads	virtTotSplitReads	splitInserts	geneA	chrA	senseA	posA	geneB	chrB	senseB	posB	dirA	dirB	dirAB_matepair	cosmicA+B
TCGA-06-6390	10	9	19	8	FGFR3	chr4	+	1778521	TACC3	chr4	+	1708787	1	1	0	2803
TCGA-12-0826	5	6	11	5	FGFR3	chr4	+	1778502	TACC3	chr4	+	1707185	0	0	1	2803
TCGA-19-5958	3	0	3	2	FGFR3	chr4	+	1778539	TACC3	chr4	+	1707203	0	1	1	2803
TCGA-27-1835	11	1	12	4	FGFR3	chr4	+	1778595	TACC3	chr4	+	1709397	0	0	1	2803
TCGA-12-0820	7	2	9	4	FGFR3	chr4	+	1779184	PRKG2	chr4	-	82338347	1	1	0	2805
TCGA-12-1088	3	1	4	4	ABL1	chr9	+	132597569	TNFRSF10B	chr8	-	22936252	0	0	1	892
TCGA-06-1802	7	1	8	8	ADAM12	chr10	-	127698245	PTPRD	chr9	-	8596127	0	0	1	54
TCGA-06-1801	7	0	7	5	HIP1	chr7	-	75010010	PTPRD	chr9	-	9387093	1	0	0	52
TCGA-12-1088	3	0	3	3	KIDINS220	chr2	-	8886300	PPP1R3A	chr7	-	113305567	0	0	1	45
TCGA-12-1088	37	1	38	10	KIDINS220	chr2	-	8887075	PPP1R3A	chr7	-	113305191	1	1	0	45
TCGA-32-2491	2	17	19	6	ODZ1	chrX	-	123342503	STAG2	chrX	+	123019118	0	1	0	36
TCGA-32-2491	11	1	12	10	ODZ1	chrX	-	123526882	SASH3	chrX	+	128749198	1	0	0	34
TCGA-12-0829	24	0	24	13	LRRK2	chr12	+	39032542	VSNL1	chr2	+	17630556	1	1	0	32
TCGA-12-0829	25	1	26	13	LRRK2	chr12	+	38975444	VSNL1	chr2	+	17639377	1	0	0	32
TCGA-12-0829	87	16	103	58	LRRK2	chr12	+	38975652	VSNL1	chr2	+	17639552	0	1	1	32
TCGA-19-0957	3	2	5	6	NUDT19	chr19	+	37891921	ODZ1	chrX	-	123925223	1	0	0	32
TCGA-12-1088	12	1	13	5	GLI3	chr7	-	42031380	RIMBP2	chr12	-	129517282	1	0	0	31
TCGA-12-1088	5	0	5	1	GLI3	chr7	-	42031574	RIMBP2	chr12	-	129517455	0	1	1	31
TCGA-12-1089	10	0	10	5	AHNAK	chr11	-	62056459	C21orf29	chr21	-	44923276	0	1	1	30
TCGA-06-1801	27	1	28	12	CROCC	chr1	+	17171362	CSMD2	chr1	-	34381139	1	0	0	29
TCGA-12-1089	12	1	13	6	CLK3	chr15	+	72705401	LRP1	chr12	+	55880002	0	1	1	28
TCGA-12-1089	14	2	16	8	CLK3	chr15	+	72705248	LRP1	chr12	+	55879646	1	0	0	28
TCGA-12-1089	42	5	47	24	LAMA2	chr6	+	129836071	PDE10A	chr6	-	165858426	1	0	0	28
TCGA-06-1802	48	9	57	27	LAMA2	chr6	+	129483265	SEC14L3	chr22	-	29193005	1	1	0	27
TCGA-19-0957	4	16	20	6	CSMD2	chr1	-	34115076	MDH2	chr7	+	75525221	0	0	1	27
TCGA-06-1801	21	1	22	4	FAM192A	chr16	-	55757701	LRP1	chr12	+	55858598	0	0	1	26
TCGA-12-1089	27	0	27	2	FGFR4	chr5	+	176447670	LILRB1	chr19	+	59840807	1	0	0	25
TCGA-19-0957	0	1	1	4	EML1	chr14	+	99349006	NRXN3	chr14	+	79233969	1	0	0	24
TCGA-06-1801	19	133	152	51	NHSL2	chrX	+	71082676	TAF1	chrX	+	70520522	1	0	0	22
TCGA-06-1801	51	3	54	8	NHSL2	chrX	+	71083319	TAF1	chrX	+	70521607	1	0	1	22
TCGA-12-1089	9	0	9	4	CACNA1C	chr12	+	2325330	ITGAV	chr2	+	187195411	0	0	1	22
TCGA-19-0957	8	1	9	6	CDH11	chr16	-	63579650	RERE	chr1	-	8588774	0	0	1	22
TCGA-12-0829	12	3	15	4	ENTPD2	chr9	-	139062591	FREM2	chr13	+	38318644	1	1	0	21
TCGA-12-0829	2	3	5	1	EFS	chr14	-	22896776	NRXN3	chr14	+	78678529	1	0	1	21

Table 2 continuation

TCGA-12-0829	56	6	62	14	DIS3L	chr15	+	64377566	GLI3	chr7	-	42032535	1	0
TCGA-12-0829	8	2	10	3	EFS	chr14	-	22896431	NRXN3	chr14	+	78678139	1	0
TCGA-12-0829	9	65	74	37	DIS3L	chr15	+	64377398	GLI3	chr7	-	42032341	1	0
TCGA-27-1835	14	0	14	4	FAM19A2	chr12	-	60707200	GLI1	chr12	+	56146523	0	0
TCGA-06-1801	20	0	20	2	FREM2	chr13	+	38163882	RALYL	chr8	+	85785432	1	1
TCGA-12-0827	2	0	2	2	ABCC12	chr16	-	46722685	FGFR4	chr5	+	176457194	1	1
TCGA-12-0829	35	0	35	7	ANXA7	chr10	-	74808655	CACNA1C	chr12	+	2458351	1	1
TCGA-06-2559	60	37	97	1	PLEKHM3	chr2	-	208426920	PTPRS	chr19	-	5222592	0	0
TCGA-12-1088	2	0	2	2	PLCL1	chr2	+	198630224	TACC2	chr10	+	123987513	1	1
TCGA-06-1801	10	0	10	4	FGFR4	chr5	+	176450528	WISP2	chr20	+	42782576	0	1
TCGA-06-1802	15	0	15	2	PDHA2	chr4	+	96980717	PDZRN4	chr12	+	39959553	0	1
TCGA-06-1802	4	0	4	2	PDHA2	chr4	+	96980509	PDZRN4	chr12	+	39959384	1	0
TCGA-06-6390	53	0	53	18	GPR182	chr12	+	55675639	PDZRN4	chr12	+	39957003	1	0
TCGA-12-0829	1121	252	1373	602	ADCY8	chr8	-	131886108	SSX3	chrX	-	48091929	0	0
TCGA-12-0829	14	8	22	3	ADCY8	chr8	-	131886506	SSX3	chrX	-	48091719	1	1
TCGA-12-0829	9	42	51	18	ADAM12	chr10	-	127733231	DAPK1	chr9	+	89454764	0	1
TCGA-12-3653	22	0	22	10	JOSD2	chr19	-	55705579	PTPRS	chr19	-	5245999	0	1
TCGA-12-0829	100	0	100	20	COL14A1	chr8	+	121370990	MMP12	chr11	-	102242881	1	0
TCGA-12-0829	152	0	152	24	COL14A1	chr8	+	121371195	MMP12	chr11	-	102242953	0	1
TCGA-06-1802	11	47	58	19	MUSK	chr9	+	112509906	SYNPO2	chr4	+	120172123	0	0
TCGA-06-1805	6	4	10	6	COL14A1	chr8	+	121332080	NCRNA0015	chr21	-	18174873	1	1
TCGA-12-0822	37	0	37	3	C7orf44	chr7	-	43683128	TACC2	chr10	+	123835337	1	0
TCGA-12-0829	0	2	2	365	GSTA3	chr6	-	52878492	TACC2	chr10	+	123884543	0	1
TCGA-12-0829	124	16	140	51	GSTA3	chr6	-	52878680	TACC2	chr10	+	123884705	0	1
TCGA-12-0829	21	7	28	10	HIP1	chr7	-	75022909	MASP1	chr3	-	188452372	0	0
TCGA-12-0829	268	123	391	242	HIP1	chr7	-	75022741	MASP1	chr3	-	188452581	1	1
TCGA-12-0829	36	641	677	365	GSTA3	chr6	-	52878496	TACC2	chr10	+	123884531	0	1
TCGA-12-1088	10	1	11	3	CAMTA1	chr1	+	7710762	TMPRSS3	chr21	-	42665918	0	1
TCGA-12-1088	65	0	65	6	ADCY10	chr1	-	166139873	DUSP27	chr1	+	165351555	0	0
TCGA-12-1088	8	1	9	4	CAMTA1	chr1	+	7714539	TMPRSS3	chr21	-	42666044	1	0
TCGA-27-1835	83	1	84	22	CMYA5	chr5	+	79120729	SRRM1	chr1	+	24870899	0	0
TCGA-06-1801	0	43	43	31	CAMTA1	chr1	+	7264935	GDPD2	chrX	+	69563759	0	1
TCGA-06-1801	13	41	54	31	CAMTA1	chr1	+	7265429	GDPD2	chrX	+	69563431	1	0
TCGA-06-1801	24	66	90	61	CAMTA1	chr1	+	7265556	GDPD2	chrX	+	69563762	0	1

Table 2 continuation

TCGA-12-0829	2	0	2	3	CCDC147	chr10	+	106165013	ISX	chr22	+	33795708	0	1
TCGA-12-1088	7	1	8	5	CMYA5	chr5	+	79045621	STK24	chr13	-	97969547	1	0
TCGA-06-1801	7	1	8	4	DEPDC5	chr22	+	30619774	ROBO1	chr3	-	79802538	0	1
TCGA-12-0820	110	20	130	23	ABCA13	chr7	+	48597322	NHSL2	chrX	+	71077547	1	0
TCGA-12-0820	29	4	33	3	ABCA13	chr7	+	48597477	NHSL2	chrX	+	71077690	0	1
TCGA-12-0829	46	2	48	4	LIN9	chr1	-	224536835	NCOR1	chr17	-	15883585	0	0
TCGA-12-3644	3	0	3	1	EFHC1	chr6	+	52432073	LRBA	chr4	-	151418615	1	0
TCGA-12-3644	3	10	13	3	EFHC1	chr6	+	52431890	LRBA	chr4	-	151418438	1	0
TCGA-19-5958	6	6	12	7	DEPDC5	chr22	+	30504095	SLC5A4	chr22	-	30974671	0	1
TCGA-06-1801	4	4	8	5	KCND3	chr1	-	112227957	LY75	chr2	-	160443238	1	1
TCGA-12-0820	26	1	27	2	BBX	chr3	+	108997451	CUL3	chr2	-	225108623	0	0
TCGA-12-0828	8	67	75	31	ADCY2	chr5	+	7558840	SDAD1	chr4	-	77096208	1	1
TCGA-12-0829	13	21	34	16	AGBL4	chr1	-	48902776	NUP188	chr9	+	130808425	0	0
TCGA-12-0829	64	308	372	197	EYS	chr6	-	64513356	IL1RN	chr2	+	113603712	0	1
TCGA-12-0829	7	25	32	11	AGBL4	chr1	-	48902600	NUP188	chr9	+	130808628	1	1
TCGA-12-0829	9	0	9	1	LRBA	chr4	-	151790893	PSEN1	chr14	+	72707609	1	0
TCGA-12-1093	65	4	69	21	OSBPL10	chr3	-	31687272	TRAPPC9	chr8	-	140828099	1	1
TCGA-12-1600	9	0	9	5	5-Sep	chr22	+	18088018	NCOR1	chr17	-	15915170	0	1
TCGA-19-0957	18	1	19	7	ADCY10	chr1	-	166060645	AKT3	chr1	-	241743142	0	1
TCGA-19-0957	34	2	36	11	ADCY10	chr1	-	166060502	AKT3	chr1	-	241742588	1	0
TCGA-12-0822	0	1	1	16	ITGB2	chr21	-	45147805	SH3RF3	chr2	+	109430489	0	1
TCGA-12-0822	6	2	8	1	ITGB2	chr21	-	45147994	SH3RF3	chr2	+	109430669	0	1
TCGA-12-0827	25	3	28	8	CUL3	chr2	-	225126210	LY75	chr2	-	160455052	1	0
TCGA-12-0828	7	2	9	4	FH	chr1	-	239743589	SRGAP1	chr12	+	62723692	0	1
TCGA-12-0829	24	0	24	9	ITGA9	chr3	+	37712050	SNX5	chr20	-	17885523	0	1
TCGA-12-1089	17	2	19	5	ABCC1	chr16	+	16077635	RNF216	chr7	-	5692038	1	1
TCGA-12-1089	6	0	6	8	CAMSAP1	chr9	-	137867066	NCF2	chr1	-	181799323	1	0
TCGA-19-0957	16	0	16	4	CCDC147	chr10	+	106114657	STK4	chr20	+	43111359	0	0
TCGA-06-1801	5	33	38	18	AP4S1	chr14	+	30611930	EYS	chr6	-	64770011	1	0
TCGA-06-1805	3	14	17	9	CUL3	chr2	-	225064315	SLC44A2	chr19	+	10608393	1	0
TCGA-12-0829	14	27	41	23	ADCY2	chr5	+	7798046	C14orf174	chr14	+	76914809	1	0
TCGA-12-0829	59	7	66	18	NR3C1	chr5	-	142760085	SORCS2	chr4	+	7354165	0	0
TCGA-12-0829	9	40	49	28	ADCY2	chr5	+	7798641	C14orf174	chr14	+	76915034	0	1
TCGA-12-1093	20	0	20	4	GAPVD1	chr9	+	127104266	MAPKAP1	chr9	-	127490362	1	0

Table 2 continuation

TCGA-12-1600	7	0	7	4	CILP	chr15	-	63283865	PARP16	chr15	-	63350048	0	1	1	10
TCGA-19-0957	13	3	16	6	AQP2	chr12	+	48635567	CDH4	chr20	+	59413648	0	1	1	10
TCGA-19-0957	6	0	6	1	AQP2	chr12	+	48635406	CDH4	chr20	+	59413468	1	0	0	10
TCGA-06-0166	2	0	2	3	CCDC158	chr4	-	77541796	SNX5	chr20	-	17885346	0	1	1	9
TCGA-06-1802	30	0	30	9	RANBP2	chr2	+	108758804	SATB2	chr2	-	199895572	0	1	1	9
TCGA-06-1805	4	0	4	3	C2CD3	chr11	-	73430819	XRRA1	chr11	-	74309669	1	1	0	9
TCGA-06-1805	6	1	7	5	NEUROG1	chr5	-	134898853	PRKCH	chr14	+	61027580	1	1	0	9
TCGA-12-0820	27	0	27	2	RANBP2	chr2	+	108749908	TTC27	chr2	+	32839367	0	1	1	9
TCGA-12-0820	58	7	65	15	RANBP2	chr2	+	108749412	TTC27	chr2	+	32837790	1	0	0	9
TCGA-12-0829	6	128	134	35	C2CD3	chr11	-	73529639	CAPZB	chr1	-	19556435	0	0	1	9
TCGA-12-0829	84	443	527	227	C2CD3	chr11	-	73529293	CAPZB	chr1	-	19556627	1	1	0	9
TCGA-12-1088	10	0	10	2	PACSIN1	chr6	+	34589431	TNC	chr9	-	116884742	0	1	0	9
TCGA-12-1088	12	0	12	2	PACSIN1	chr6	+	34589619	TNC	chr9	-	116884958	1	0	1	9
TCGA-19-0957	34	19	53	17	PRKCH	chr14	+	61032978	ZFAND3	chr6	+	38228111	1	0	1	9
TCGA-19-0957	7	1	8	7	MAPKAP1	chr9	-	127348507	SLC9A1	chr1	-	27302334	0	1	0	9
TCGA-19-0957	8	39	47	21	PRKCH	chr14	+	61032774	ZFAND3	chr6	+	38227949	1	0	0	9
TCGA-06-1801	5	11	16	4	MAOA	chrX	+	43486192	SH3RF3	chr2	+	109237058	1	0	1	8
TCGA-06-1802	10	12	22	12	DNM1L	chr12	+	32736794	SYNPO2	chr4	+	120172271	0	1	1	8
TCGA-06-1802	18	42	60	24	MUC4	chr3	-	196982875	SMOC2	chr6	+	168676813	1	0	1	8
TCGA-12-0829	6	0	6	6	ATXN1	chr6	-	16669201	CACNA1G	chr17	+	46004995	0	0	1	8
TCGA-12-0829	7	0	7	3	ATP6V0D2	chr8	+	87186716	RERE	chr1	-	8336574	1	1	0	8
TCGA-12-1088	11	1	12	6	BCAS3	chr17	+	56321892	CACNA1G	chr17	+	46010698	1	1	0	8
TCGA-12-1088	15	2	17	5	ABCC1	chr16	+	16135771	AGBL4	chr1	-	49315120	1	0	0	8
TCGA-12-1088	17	3	20	4	MST1R	chr3	-	49910627	WDFY1	chr2	-	224512774	1	0	0	8
TCGA-12-1088	4	0	4	2	FBXL4	chr6	-	99431443	SYNPO2	chr4	+	120172560	0	0	1	8
TCGA-12-1092	39	4	43	12	CNTN2	chr1	+	203302926	DNAJC6	chr1	+	65591195	0	0	1	8
TCGA-12-1598	4	0	4	5	MPP1	chrX	-	153673715	SRGAP1	chr12	+	62777947	0	0	1	8
TCGA-19-1786	5	19	24	7	ATP5B	chr12	-	55320148	USP48	chr1	-	21920103	1	0	1	8
TCGA-19-2621	21	0	21	3	BCAS3	chr17	+	56731673	TTYH1	chr19	+	59638801	1	1	0	8
TCGA-06-1801	15	0	15	9	C15orf23	chr15	+	38469150	DMD	chrX	-	32092185	0	1	1	7
TCGA-06-1805	6	3	9	5	FAM19A2	chr12	-	60547321	POLM	chr7	-	44082653	0	1	0	7
TCGA-12-0829	13	84	97	44	ATP5B	chr12	-	55318484	PRC1	chr15	-	89330475	1	0	1	7
TCGA-12-0829	158	207	365	44	ATP5B	chr12	-	55320850	PRC1	chr15	-	89334458	1	1	0	7
TCGA-12-0829	2	1	3	2	DDI2	chr1	+	15825507	KIDINS220	chr2	-	8805399	0	1	0	7

Table 2: continuation

TCGA-12-0829	25	6	31	44	ATP5B	chr12 -	55321832	PRC1	chr15 -	89335627	1	0	1	7
TCGA-12-0829	34	21	55	44	ATP5B	chr12 -	55321200	PRC1	chr15 -	89335044	0	1	0	7
TCGA-12-0829	44	6	50	4	ABCC6	chr16 -	16204784	SUMF1	chr3 -	4470138	1	0	0	7
TCGA-12-0829	53	28	81	35	DDI2	chr1 +	15825941	KIDINS220	chr2 -	8805580	1	0	1	7
TCGA-12-0829	9	0	9	5	DMD	chrX -	32013100	N4BP2L2	chr13 -	32008512	0	0	1	7
TCGA-12-1092	9	0	9	4	LRRC4B	chr19 -	55754780	NR3C1	chr5 -	142660156	0	1	0	7
TCGA-19-2621	3	22	25	11	PCDH12	chr5 -	141309153	SLC36A2	chr5 -	150679274	1	1	0	7
TCGA-06-1802	8	4	12	6	BAHD1	chr15 +	38539023	OSBPL10	chr3 -	31729622	0	1	1	6
TCGA-12-0828	11	0	11	1	PLOD3	chr7 -	100646340	VSNL1	chr2 +	17638618	1	1	0	6
TCGA-12-0828	40	9	49	20	PLOD3	chr7 -	100646511	VSNL1	chr2 +	17637955	0	0	1	6
TCGA-12-0829	16	2	18	9	C21orf29	chr21 -	44922864	MYT1	chr20 +	62300829	0	0	1	6
TCGA-12-0829	196	0	196	37	IGFBP3	chr7 -	45922866	SMOC2	chr6 +	168722450	0	0	1	6
TCGA-12-0829	5	1	6	1	FAM168A	chr11 -	72839771	NCF2	chr1 -	181826115	1	0	1	6
TCGA-12-0829	5	18	23	9	FAM168A	chr11 -	72839534	NCF2	chr1 -	181825930	1	0	0	6
TCGA-12-1089	20	0	20	2	SLC44A2	chr19 +	10602997	XRCC4	chr5 +	82430803	1	0	0	6
TCGA-19-0957	17	1	18	4	PAX3	chr2 -	222778052	WDFY1	chr2 -	224453159	1	0	1	6
TCGA-06-1801	5	0	5	1	CAP2	chr6 +	17571234	DNAJC6	chr1 +	65602700	1	0	0	5
TCGA-06-1801	6	32	38	15	CAP2	chr6 +	17571666	DNAJC6	chr1 +	65603089	0	1	1	5
TCGA-06-1805	3	0	3	8	PLCL1	chr2 +	198578552	SURF6	chr9 -	135188818	0	1	0	5
TCGA-06-1805	7	2	9	4	PLCL1	chr2 +	198578671	SURF6	chr9 -	135189294	1	0	1	5
TCGA-12-0822	17	4	21	4	TAAR6	chr6 +	132933266	TTYH1	chr19 +	59629451	0	1	1	5
TCGA-12-0828	17	0	17	8	AQP2	chr12 +	48634800	ECE1	chr1 -	21515240	0	1	1	5
TCGA-12-0828	7	0	7	1	AQP2	chr12 +	48634610	ECE1	chr1 -	21515033	1	0	0	5
TCGA-12-0829	12	0	12	7	CACNA1G	chr17 +	46039372	CNTNAP4	chr16 +	74873868	0	1	1	5
TCGA-19-0957	4	0	4	3	PCDH12	chr5 -	141316624	SH3BP5	chr3 -	15315567	0	0	1	5
TCGA-19-0957	8	2	10	2	PCDH12	chr5 -	141316405	SH3BP5	chr3 -	15315731	1	1	0	5
TCGA-06-1801	13	1	14	1	ABCC6	chr16 -	16205051	CMTM7	chr3 +	32443880	0	1	1	4
TCGA-06-1801	33	7	40	4	ABCC6	chr16 -	16204860	CMTM7	chr3 +	32443722	1	0	0	4
TCGA-06-1805	10	3	13	6	AGBL4	chr1 -	49449813	NOX4	chr11 -	88714996	0	0	1	4
TCGA-12-0829	11	0	11	4	FAM160A1	chr4 +	152595916	LY75	chr2 -	160440194	0	1	0	4
TCGA-12-0829	17	1	18	5	FAM160A1	chr4 +	152596097	LY75	chr2 -	160440376	1	0	1	4
TCGA-12-0829	487	83	570	249	CORO7	chr16 -	4375428	DYRK3	chr1 +	204876154	0	0	1	4
TCGA-12-1088	2	12	14	4	FAM172A	chr5 -	93052315	TRIOBP	chr22 +	36427382	0	1	1	4
TCGA-06-1801	30	7	37	18	DEPDC7	chr11 +	33003811	EIF2C2	chr8 -	141618836	1	0	0	3

CA 02888001 2015-01-23

Table 2 continuation

TCGA-06-1801	40	28	68	33	MAP7	chr6	-	136728609	SH3RF3	chr2	+	109392677	0	0
TCGA-12-1093	6	15	21	8	CORO7	chr16	-	4398302	PLEK2	chr14	-	66934201	0	1
TCGA-12-3644	33	0	33	4	EDA	chrX	+	69073054	SSX3	chrX	-	48094443	1	1
TCGA-12-3644	37	15	52	17	C15orf33	chr15	-	47424122	PARP16	chr15	-	63350289	0	0
TCGA-19-1791	14	3	17	8	PSEN1	chr14	+	72748293	ZNF410	chr14	+	73431112	0	0
TCGA-06-1802	35	2	37	17	CELF2	chr10	+	11352537	PLA2G2F	chr1	+	20348173	1	1
TCGA-06-1802	63	26	89	25	CELF2	chr10	+	11352765	PLA2G2F	chr1	+	20347997	0	0
TCGA-06-1802	8	2	10	8	LCLAT1	chr2	+	30535977	PACSIN1	chr6	+	34576195	1	0
TCGA-06-2562	6	0	6	4	SNTA1	chr20	-	31473415	TMEM80	chr11	+	689744	0	0
TCGA-12-0829	16	1	17	4	LASS6	chr2	+	169045211	NKAIN2	chr6	+	125021252	0	0
TCGA-12-0829	7	0	7	2	LASS6	chr2	+	169045333	NKAIN2	chr6	+	125021072	1	1
TCGA-14-0813	339	39	378	5	SNTA1	chr20	-	31481069	TMEM80	chr11	+	686739	0	0
TCGA-12-0820	49	3	52	11	CAMKK1	chr17	-	3712344	FAM184B	chr4	-	17271273	0	1
TCGA-12-0826	8	18	26	13	CELF2	chr10	+	11406463	NME4	chr16	+	389427	1	1
TCGA-12-1089	17	4	21	10	C6orf170	chr6	-	121478035	NKAIN2	chr6	+	125083380	1	0
TCGA-12-1600	19	0	19	3	ATP6AP1L	chr5	+	81649744	FAM172A	chr5	-	93336459	1	0
TCGA-12-1600	5	35	40	6	ATP6AP1L	chr5	+	81649902	FAM172A	chr5	-	93336676	1	0
TCGA-19-1790	4	0	4	4	ARMC6	chr19	+	19026932	FAM184B	chr4	-	17391210	1	0
TCGA-06-1802	12	0	12	8	EIF2C2	chr8	-	141648334	TNFRSF10B	chr8	-	22940680	0	0
TCGA-14-0781	22	2	24	8	FAM160A1	chr4	+	152584637	UNC93B1	chr11	-	67523253	1	0

Table 3 Recurrent gene fusion pairs from Exome-fuse analysis of 84 GBM from TCGA.

Sample	gene A	gene B	Sample	gene A	gene B
TCGA-12-0820	ABCA13	NHSL2	TCGA-12-0820	CAMKK1	FAM184B
TCGA-12-1089	ABCC1	RNF216	TCGA-12-1089	CAMSAP1	NCF2
TCGA-12-1088	ABCC1	AGBL4	TCGA-12-1088	CAMTA1	TMPRSS3
TCGA-12-0827	ABCC12	FGFR4	TCGA-06-1801	CAMTA1	GDPD2
TCGA-12-0829	ABCC6	SUMF1	TCGA-06-1801	CAP2	DNAJC6
TCGA-06-1801	ABCC6	CMTM7	TCGA-19-0957	CCDC147	STK4
TCGA-12-1088	ABL1	TNFRSF10B	TCGA-12-0829	CCDC147	ISX
TCGA-06-1802	ADAM12	PTPRD	TCGA-06-0166	CCDC158	SNX5
TCGA-12-0829	ADAM12	DAPK1	TCGA-19-0957	CDH11	RERE
TCGA-12-1088	ADCY10	DUSP27	TCGA-06-1802	CELF2	PLA2G2F
TCGA-19-0957	ADCY10	AKT3	TCGA-12-0826	CELF2	NME4
TCGA-12-0828	ADCY2	SDAD1	TCGA-12-1600	CILP	PARP16
TCGA-12-0829	ADCY2	C14orf174	TCGA-12-1089	CLK3	LRP1
TCGA-12-0829	ADCY8	SSX3	TCGA-12-1088	CMYA5	STK24
TCGA-12-0829	AGBL4	NUP188	TCGA-27-1835	CMYA5	SRRM1
TCGA-06-1805	AGBL4	NOX4	TCGA-12-1092	CNTN2	DNAJC6
TCGA-12-1089	AHNAK	C21orf29	TCGA-06-1805	COL14A1	NCRNA00157
TCGA-12-0829	ANXA7	CACNA1C	TCGA-12-0829	COL14A1	MMP12
TCGA-06-1801	AP4S1	EYS	TCGA-12-1093	CORO7	PLEK2
TCGA-12-0828	AQP2	ECE1	TCGA-12-0829	CORO7	DYRK3
TCGA-19-0957	AQP2	CDH4	TCGA-06-1801	CROCC	CSMD2
TCGA-19-1790	ARMC6	FAM184B	TCGA-19-0957	CSMD2	MDH2
TCGA-19-1786	ATP5B	USP48	TCGA-06-1805	CUL3	SLC44A2
TCGA-12-0829	ATP5B	PRC1	TCGA-12-0827	CUL3	LY75
TCGA-12-1600	ATP6AP1L	FAM172A	TCGA-12-0829	DDI2	KIDINS220
TCGA-12-0829	ATP6V0D2	RERE	TCGA-19-5958	DEPDC5	SLC5A4
TCGA-12-0829	ATXN1	CACNA1G	TCGA-06-1801	DEPDC5	ROBO1
TCGA-06-1802	BAHD1	OSBPL10	TCGA-06-1801	DEPDC7	EIF2C2
TCGA-12-0820	BBX	CUL3	TCGA-12-0829	DIS3L	GLI3
TCGA-19-2621	BCAS3	TTYH1	TCGA-12-0829	DMD	N4BP2L2
TCGA-12-1088	BCAS3	CACNA1G	TCGA-06-1802	DNM1L	SYNPO2
TCGA-06-1801	C15orf23	DMD	TCGA-12-3644	EDA	SSX3
TCGA-12-3644	C15orf33	PARP16	TCGA-12-3644	EFHC1	LRBA
TCGA-12-0829	C21orf29	MYT1	TCGA-12-0829	EFS	NRXN3
TCGA-06-1805	C2CD3	XRRA1	TCGA-06-1802	EIF2C2	TNFRSF10B
TCGA-12-0829	C2CD3	CAPZB	TCGA-19-0957	EML1	NRXN3
TCGA-12-1089	C6orf170	NKAIN2	TCGA-12-0829	ENTPD2	FREM2
TCGA-12-0822	C7orf44	TACC2	TCGA-12-0829	EYS	IL1RN
TCGA-12-1089	CACNA1C	ITGAV	TCGA-14-0781	FAM160A1	UNC93B1
TCGA-12-0829	CACNA1G	CNTNAP4	TCGA-12-0829	FAM160A1	LY75

[00261] **Table 3** above shows recurrent gene fusion pairs from Exome-fuse analysis of 84 GBM from TCGA. Fusion candidates have been nominated if they have at least two split inserts and at least two split reads. To further filter the list on recurrence, any fusion candidate was kept in which one of the genes is involved in at least two fusions across

different samples.

[00262] To experimentally validate the computational predictions that emerged from TX-Fuse, the PCR products spanning the fusion breakpoint were sequenced and validated each of the five in-frame fusion predictions (**Figures 1** and **9**). In **Figure 1B**, the prediction is shown and in **Figure 1C**, the cDNA sequence validation for the fusion with the highest read support involving FGFR3 fused in-frame with TACC3 in GSC-1123 is shown. The same FGFR3-TACC3 fusion transcript was also detected in the primary GBM-1123 tumor specimen from which the GSC-1123 culture was established (**Figure 1C**). The amplified cDNA contained an open reading frame for a protein of 1,048 amino acids resulting from the fusion of a FGFR3 amino-terminal portion of residues 1-758 with a TACC3 carboxy-terminal portion of residues 549-838 (**Figure 1D**). FGFR3 is a member of the FGFR receptor tyrosine kinase (TK) family that transduces intracellular signals after binding to FGF ligands (Turner and Grose, 2010). TACC3 belongs to the evolutionarily conserved TACC gene family, which also includes TACC1 and TACC2. The distinctive feature of TACC proteins is the presence of a coiled-coil domain at the C-terminus, known as the TACC domain. Through the TACC domain, TACC proteins localize to the mitotic spindle during metaphase and stabilize the microtubule spindle network (Hood and Royle, 2011; Peset and Vernos, 2008). In the predicted fusion protein the intracellular TK domain of FGFR3 is fused upstream of the TACC domain of TACC3 (**Figure 1D**).

[00263] Exon-specific gene expression analysis from the RNA-seq coverage in GSC-1123 demonstrated that the FGFR3 and TACC3 exons implicated in the fusion are highly overexpressed compared with the mRNA sequences not included in the fusion event (**Figure 10A**). Quantitative RT-PCR showed that the expression of the fused FGFR3-TACC3 exons is significantly higher in GSC-1123 than other GSCs and the normal brain (80 to 130-fold, **Figure 10B**). Without being bound by theory, functionally significant genetic rearrangements may result in marked overexpression (outlier) of the genes implicated in the fusion events (Tomlins et al., 2007; Tomlins et al., 2005). The FGFR3-TACC3 fusion protein was also abundantly expressed in GSC-1123 and in the primary tumor GBM-1123, as shown by Western blot and immunohistochemistry (**Figures 10C and 10D**). On a Western Blot, the FGFR3-TACC3 fusion protein migrated at a size of ~150 kD and immunoprecipitation followed by mass spectrometry revealed the presence of FGFR3 and TACC3 peptides consistent with the cDNA translation prediction (**Figure 10E**). Using PCR, the genomic

breakpoint coordinates were mapped to chromosome 4 (#1,808,966 for FGFR3 and #1,737,080 for TACC3, genome build GRCh37/hg19) falling within FGFR3 exon 17 and TACC3 intron 7, which gives rise to a transcript in which the 5' FGFR3 exon 16 is spliced to the 3' TACC3 exon 8. The DNA junctions of FGFR3 and TACC3 show microhomology within a 10-base region, an observation consistent with results previously reported for other chromosomal rearrangements in human cancer (Bass et al., 2011; Stephens et al., 2009) (**Figure 1E**).

[00264] The experimental validation of the inferred genomic fusions was focused on FGFR3-TACC3. Exome-Fuse identified FGFR3-TACC3 gene fusions in four GBM samples with breakpoints spanning invariably within intron 16 of FGFR3 (which is downstream to the coding region for the TK domain) and intron 7-10 of TACC3 (which is upstream to the TACC domain) (**Figure 2A**, Tables 4 and 5). Among the four positive TCGA GBM specimens, two were available from TCGA centers for molecular analysis (TCGA-27-1835 and TCGA-06-6390) and, by Sanger sequencing, each of them were confirmed to carry an in-frame fusion transcript that is consistent with the predicted genomic breakpoints (**Figures 2B** and **2C**). Thus, the frames of the FGFR3-TACC3 fusion proteins invariably result in juxtaposing the TK domain of FGFR3 upstream of the TACC domain of TACC3. Consistent with the abundant expression of FGFR3-TACC3 in GSC-1123 and GBM-1123, the mRNA expression analysis of the TCGA tumors revealed that the four FGFR3-TACC3-positive GBM display marked co-outlier expression of FGFR3 and TACC3 (**Figure 2D**). Recurrent gene fusions can be associated with local copy number variations (CNV) of the breakpoint regions (Wang et al., 2009). Accordingly, the analysis of SNP arrays in the TCGA dataset revealed the presence of microamplification events of the FGFR3 and TACC3 genes in all four FGFR3-TACC3-positive GBM (**Figure 2E**).

Table 4: List of split inserts supporting the identification of FGFR3-TACC3 fusion genes in four GBM samples from the ATLAS-TCGA exome collection

TCGAsampleID	gene1	gene1 length	readID	%identity	length	mismatch	gap	read start	read end	hg18 genome	hg18 genome	e-value	bit score	read 1 fasta
TCGA-02-6690	FGFR3	76	C01PRACXX:110628:8:1301:16934:116558	100	76	0	0	1	76	1775372	1775447	8E-40	151	GTGCTGGCATGCCGCGCCTCTCCAGAGGCCCACC TTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCT TACCGTG
TCGA-02-6690	FGFR3	76	C01RDACXX:110628:3:2925:4672:47008	99.86	76	1	0	1	76	1775364	1775439	2E-37	143	ATGCGGGAGTGCTGGCATGACCGCCCTCCCAGAG GCCCACCTTCAAGCAGCTGGTGGAGGACCTGGACC GTGTCC
TCGA-02-6690	FGFR3	76	D03U9ACXX:110625:8:1203:16178:182312	100	76	0	0	1	76	1775413	1775488	8E-40	151	AGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTG ACGTCCAGCGACGTGAGTGCTGGCTTTGGCCTGGT GCCACC
TCGA-02-6690	TACC3	76	C01PRACXX:110628:2:1102:13552:102312	100	76	0	0	1	76	1705818	1705843	8E-40	151	GTCCA AACCAAAGGCTC4GACCCCCAGGA4TAGAA4ATATA GGCCCT74AA4CAACTCGTTCCCTCAG4CCAC4CAC AAG4
TCGA-02-6690	TACC3	76	C01PRACXX:110628:5:2308:8515:60354	100	76	0	0	1	76	1705856	1705881	8E-40	151	TCAAGGACTT4CAGGAATGTCCA3TGCTCCCAAGAA ATCGA4CTCC4CAAGCTTGGCTTCCCGCGC4CGTC
TCGA-02-6690	TACC3	76	C01RDACXX:110628:8:1305:16343:57213	98.68	76	1	0	1	76	1705886	1705760	2E-37	142	CTGAG GCACTT4CAGGA4TGTCCAGTGCTCCCAA3AAATCG
TCGA-02-6690	TACC3	76	D03U9ACXX:110625:2:1202:12578:80251	100	75	0	0	1	75	1705861	1705787	3E-32	142	AACTCCACAAGCTTGGCTTCCCGCGGGACGTCCTCA GGGAT
TCGA-02-6690	TACC3	76	C01C59AAXX:100217:4:2:4279:6949	100	76	0	0	1	76	1707299	1707224	8E-40	151	CAGACCACACACAAGACACATTCAAGAGGCACTCAA GCACTT4CAGGA4TGTCCAGTGCTCCCAA3AAATCG AACTC
TCGA-02-6690	FGFR3	76	42MJNAAXX:89813:5:37:436:12508	100	76	0	0	1	76	1775346	1775421	8E-40	151	CTTACCGTGACGTCCACCGACGAGTGAGTGCTGGCTCTGG ACCTGGTGCCACCCGCCGATCTCTCTCCCCTGTCCT TTTCCT
TCGA-02-6690	FGFR3	51	61C59AAXX:100217:5:89:7737:2567	98.04	51	1	0	1	51	1778443	1778493	4E-24	93.7	TGGGAGGGTGCGGGGGGGCCCGGGGGGGGGGAGTGT GCAGGTGAGCTCCCTGGCCCTTGGCCCCCTGCCCT CTGGGGGG
TCGA-12-5958	TACC3	62	D03U24CXX:110625:4:2236:9451:114189	90.32	62	0	0	1	62	1707141	1707202	4E-17	75.8	CTGGGAATGGTGGTGTCTGGGGCAGGGTTGTGGGT GACCGGGGGTGGGAGGGTGCGGGGGACCGGGGG
TCGA-12-5958	TACC3	74	D03U24CXX:110625:1:2204:20004:21122	95.95	74	0	0	1	74	1707097	1707170	3E-33	128	GGGGAGGG
TCGA-27-1835	FGFR3	76	C00HWABXX:110325:7:2220:17850:110366	100	76	0	0	1	76	1775338	1775413	8E-40	151	AGCGCCCTGCCCGCAGGT4CATGATCATGGGGAG TGCTGGCATGCCGCGCCCTCCC4GAGGCCCACCTT CAAGCA
TCGA-27-1835	TACC3	76	C00HWAEXX:110325:7:2104:10731:61183	100	76	0	0	1	76	1709492	1709417	8E-40	151	GCCAACGCCATGCCCAGGCCGGAGAGTCCCGGGG AGGCTGCTGGTGGGCAGCTGACTGCGGGGACACT GGGTGGAA
TCGA-27-1835	TACC3	60	E02VI2ABXX:102426:2:2201:5811:24541	100	60	0	0	1	60	1709504	1709445	3E-20	112	CGCGGGGA TGCCCAGGCCGGAGAGTCCCGGGGCGGCTGCTGG GGGGGAGCTGACTGGGGGGGCACTGGGGGGGAG ACCCGGGCC

Table 4 (Cont.)

TCGAsampleID	gene2	gene2 length	readID	%identity	length	mismatch	gap	read start	read end	hg18 genome start	hg18 genome end	e-value	bit score	read2 fasta
TCGA-08-0390	TACC3	76	C01PRACXX:110828:1:1201:1994:116558	100	76	0	0	1	76	1708822	1708647	8E-40	151	TAGGCCCTTAAAACAACTCGTTCCCTCAGACCACAC ACAAGACAGTTCAAGAGGGACTCAAGGACTTACAG GAATG
TCGA-08-0390	TACC3	76	C01PRACXX:110828:3:2305:4572:47502	100	76	0	0	1	76	1708887	1708792	8E-40	151	ACTCAAGGACTTACAGGAATGTCCAGTCGTCCCAAG AAATCGAACTCCACAAGCTTGGCTTCCCGCGGACCG TCCTG
TCGA-08-0390	TACC3	76	D03U9ACXX:110828:5:1203:10178:139219	100	76	0	0	1	76	1708821	1708646	8E-40	151	AGGCCCTTAAAACAACTCGTTCCCTCAGACCACACA CAAGACAGTTCAAGAGGGACTCAAGGACTTACAGG AATGT
TCGA-08-0390	FGFR3	76	C01PRACXX:110828:2:1102:13552:129312	100	76	0	0	1	76	1778387	1778462	8E-40	151	GCCCTCCCAGAGGCCCACCTTCAAGCAGCTGGTGG AGGACCTGGACCGTGTCCTTACCGTGACGTCCACC GACGTG
TCGA-08-0390	FGFR3	76	C01PRACXX:110828:8:2308:6515:60364	100	76	0	0	1	76	1778382	1778457	8E-40	151	GCCGCGCCCTCCCAGAGGCCCACCTTCAAGCAGCT GGTGGAGGACCTGGACCGTGTCCTTACCGTGACGT CCACCG
TCGA-08-0390	FGFR3	76	C01PRACXX:110828:6:1306:16943:57213	96.05	76	3	0	1	76	1778417	1778492	1E-32	127	GGTGGAGGACCTGGACCGTGACCTTACCGTGACG TCCACCGACGGGAGTGCTGGCTCTGGCCTGGTGC CACCGCC
TCGA-08-0390	FGFR3	76	D03U9ACXX:110828:5:2102:19575:80281	100	76	0	0	1	76	1778447	1778522	8E-40	151	GACGTCCACCGAGTGAGTGCTGGCTCTGGCCTGGTGC GTGCCACCCGCCTATGCCCCTCCCCCTGCCGTCCC CGGCCAT
TCGA-08-0390	FGFR3	76	D03U9ACXX:110828:4:2202:2824:174272	100	76	0	0	1	76	1778436	1778510	8E-40	151	TGTCCTTACCGTGACGTCCACCGACGTGAGTGCTG GCTCTGGCCTGGTGCCACCCGCCTATGCCCCTCCC CCTGCC
TCGA-12-0826	FGFR3	76	61C52AAXX:100217:4:2407:8842	98.68	76	0	1	1	76	1778436	1778510	8E-37	141	AAAAGATTTAAGTTTAGATCTTTAATATACCTAGAACG GTGGCTGTAACCAGCAAGGCAGGAGCCCTTTGTGT TGG
TCGA-12-0826	TACC3	67	40MJMAAXX:90813:6:37:435:12500	98.51	67	1	0	1	67	1707636	1707569	5E-32	125	TGGGTCAAACTTGAGGTATAAGGACTGCTTCCTCAA GGCCGACTCCTTATACTGCGGACAAGAGGGCAAGT GATCA
TCGA-12-0822	TACC3	75	61C52AAXX:100217:5:96:7727:2567	97.33	75	2	0	2	76	1707306	1707232	8E-38	133	GAGTGCTGGCTCTGGCCTGGTGCCACCCGCCTATG CCCCTCCCCCTGCCGTCCCCGGCCATCCTGCCCCC CAGAGT
TCGA-19-5958	FGFR3	76	D03U9ACXX:110828:4:2206:6451:114169	98.68	76	1	0	1	76	1778462	1778537	8E-37	143	GAGTGCTGGCTCTGGCCTGGTGCCACCCGCCTATG CCCCTCCCCCTGCCGTCCCCGGCCATCCTGCCCCC CAGAGT
TCGA-19-5960	FGFR3	76	D03U9ACXX:110828:1:2204:20084:21192	100	76	0	0	1	76	1778462	1778537	2E-41	161	GCCAACCCCATGCCCAGGCCCGGAGAGTCCCGGGG AGGCTGCTGGTGGGGAGCTGACTTCGGGGACACT CGGGGGAA
TCGA-27-1835	TACC3	76	C09HWABXX:110825:7:2202:17680:192669	96.05	76	3	0	1	76	1708492	1708417	1E-32	127	CATGCGGGAGTGCTGCCATGGCCCCCCCTCCCAG CGGCCCACCTTCAAGCAGCTGGTGGGGGACCTGC ACCGTGTC
TCGA-27-1835	FGFR3	76	C09HWABXX:110825:7:2104:10731:5183	96.05	76	3	0	1	76	1778363	1778438	1E-32	127	ACGTGAGTGCTGGCTCTGGCCTGGTGCCACCCGC CTATGCCCCTCCCCCTGCCGTCCCCGGCCATCCTG CCCCCCA
TCGA-27-1836	FGFR3	76	E09V2ABXX:110408:2:2201:5831:24541	100	76	0	0	1	76	1778458	1778533	8E-40	151	CCCTCCCCAGAGGCCCACCTTCAAGCAGCTGGTGGA GGAGCTGGACCGTGTCCTTACCGTGACGTCCACCG ACGTGA
TCGA-27-1835	FGFR3	76	B097UABXX:110405:4:2102:15742:63594	100	76	0	0	1	76	1778388	1778463	8E-40	151	

Table 5: List of split reads supporting the identification of FGFR3-TACC3 fusion genes in four GBM samples from the ATLAS-TCGA exome collection

sample	genesplit1	readID	directionsplit	fig18startsplit1	fig18stopsplit1	length1	mismatch1	gap1	seqsplit
TCGA-06-6390	TACC3	D02UGACXX110625:2:1202:19579:92281	F	1778621	1778621	1	0	0	GGACTTACAGGAATGTTCAGTTGCTCCCAAGAAATCGAACTCCC ACAAGCTTGGCTTCCCGCGGAACGTCCTGAGGATA***T
TCGA-06-6390	FGFR3	D01PPACXX110628:3:11C4:10952:86371	F	1778620	1778621	2	0	0	CA***TCCCTCAGGAACGTCCGCCGGGAAGCCAAGCTTGTTGGAG TTCGATTTCTTGGGAGCACTGGACATTCCTGTAAGTC
TCGA-06-6390	FGFR3	C01PRACXX110628:6:1108:2119:92993	F	1778620	1778621	2	0	0	CA***TCCCTCAGGAACGTCCGCCGGGAAGCCAAGCTTGTTGGAG TTCGATTTCTTGGGAGCACTGGACATTCCTGTAAGTC
TCGA-06-6390	FGFR3	D02UGACXX110625:6:2304:12523:106352	F	1778620	1778621	2	0	0	CA***TCCCTCAGGAACGTCCGCCGGGAAGCCAAGCTTGTTGGAG TTCGATTTCTTGGGAGCACTGGACATTCCTGTAAGTC
TCGA-06-6390	FGFR3	C01PRACXX110628:5:2103:6915:97945	F	1778618	1778621	4	0	0	GCCA***TCCCTCAGGAACGTCCGCCGGGAAGCCAAGCTTGTTGG AGTTCGATTTCTTGGGAGCACTGGACATTCCTGTAAG
TCGA-06-6390	FGFR3	C01PRACXX110628:8:1204:10631:92926	F	1778612	1778621	10	0	0	TTGTTGGAGTTCGATTTCTTGGGAGCACTGGACATTCC TCCCCCGCCA***TCCCTCAGGAACGTCCGCCGGGAAGCCAAGC
TCGA-06-6390	FGFR3	C01PRACXX110628:5:2204:6732:191380	F	1778612	1778621	10	0	0	TTGTTGGAGTTCGATTTCTTGGGAGCACTGGACATTCC TCCCCCGCCA***TCCCTCAGGAACGTCCGCCGGGAAGCCAAGC
TCGA-06-6390	FGFR3	C01PRACXX110628:8:1308:2831:23560	F	1778611	1778621	11	0	0	GCTTGTTGGAGTTCGATTTCTTGGGAGCACTGGACATTC CCCCTCCCCGCCA***TCCCTCAGGAACGTCCGCCGGGAAGCC
TCGA-06-6390	FGFR3	C01PRACXX110628:5:2207:9589:94017	F	1778609	1778621	13	0	0	AAGCTTGTTGGAGTTCGATTTCTTGGGAGCACTGGACAT TCCCTCCCCGCCA***TCCCTCAGGAACGTCCGCCGGGAAGCC
TCGA-06-6390	TACC3	C01PRACXX110628:7:2205:11825:39734	R	1778604	1778621	21	0	0	GAAGTCCTGAGGGA***TGGCCCGGGGACGGCAGGGGCA AAGAAATCGAACTCCACAAGCTTGGCTTCCCGCGGGACGTCC
TCGA-06-6390	TACC3	C01PRACXX110628:6:1108:12159:179469	R	1778499	1778621	28	0	0	TGAGGGA***TGGCCGGGGACGGCAGGGGCAGGGGCAT AAATCGAACTCCACAAGCTTGGCTTCCCGCGGACGTCCTGA
TCGA-06-6390	TACC3	D02UGACXX110625:4:2202:12501:40288	R	1778491	1778621	21	1	0	GGGA***TGGCCGGGGGCGGCAGGGGGAGGGGCATAGG GTGGCCTGGTGCCACCCGCCTATGCCCCTCCCCCTGCCTC
TCGA-06-6390	FGFR3	C01RDACXX110628:3:1305:3044:13228	F	1778472	1778621	49	0	0	CCCGCCA***TCCCTCAGGAACGTCCGCCGGGAAGCCAA TCGTCCCGCCGGACTTCCTGATGGA***TGGCCCGGGGACGGCA
TCGA-06-6390	TACC3	D02UGACXX110625:6:2206:12523:106352	R	1778470	1778621	52	4	0	GGGGGAGGGGCATAGGCGTGTGGCACCAGGCCAGAGCC CTTCCCGCCGGACGTCCTGAGGGA***TGGCCCGGGGACGGNA
TCGA-06-6390	TACC3	C01PRACXX110628:7:2205:11825:39734	R	1778466	1778621	52	1	0	GGGGGAGGGGCATAGGCGGGGTGGCACCAGGCCAGAGCC GTGCCTGGCTCTGCCCTGGTGCCACCCGCCTATGCCCCCCTCCC
TCGA-06-6390	FGFR3	C02U9ACXX110625:7:2106:4402:172653	F	1778464	1778621	58	0	0	CCTGCCGTCCCCGGCCA***TCCCTCAGGAACGTCCGCCG GAGGGA***TGGCCGGGGGACGGCAGGGGCAGGGGCCATAGG
TCGA-06-6390	TACC3	C01PRACXX110628:8:2103:6915:17943	R	1778452	1778621	70	0	0	CGGGTGGCACCAGGCCAGAGCCAGCACTCACGTCGGGTGG
TCGA-12-0826	TACC3	61DG9AAXX100217:4:23:15135:61183	R	1778496	1778602	3	0	0	GGACAAGAAGGGCAAGTGATCAGGTCTGACTGCCATCCCCTA ACACACACAGGGGGCTAAGGGCAGGG***GAGGGGCA
TCGA-12-0826	TACC3	61DG9AAXX100217:5:107:10675:10040	R	1778495	1778602	5	0	0	GGACAAGAAGGGCAAGTGATCAGGTCTGACTGCCATCCCCTA ACACACACAGGGGGCTAAGGGCAGGG***GAGGGGCA
TCGA-12-0826	FGFR3	61DG9AAXX100217:5:108:1909:11295	F	1778494	1778602	6	0	0	ATGCCCCTC***CCCTGCCCTTAGCCCCCCTGTGTGTGTTAGG GGATGCCAGTCAGACCTGATCACTTGCCCTCTTGTC
TCGA-12-0826	FGFR3	61DG9AAXX100217:5:82:13129:10087	F	1778490	1778602	13	0	0	GCCTATGCCCCTC***CCCTGCCCTTAGCCCCCCTGTGTGTGT TAGGGCATGCCAGTCAGACCTGATCACTTGCCCTCT
TCGA-12-0826	FGFR3	42MJN4AAXX00606:3:5:23:1551:1150#0	F	1778481	1778602	22	0	0	GTGCCACCCCGCCTATGCCCCTC***CCCTGCCCTTAGCCCCC GTGTGTGTGTTTAGGGCATGCCAGTCAGACCTGATCAC
TCGA-12-0826	FGFR3	61DG9AAXX100217:4:23:14379:6949	F	1778438	1778602	66	0	0	TGACTGCCATCCCCTAACACACACAGGGGGCTAAGGGCAG GG***GAGGGGCATAGGCGGGGGGCAGCCAGGCCAGAGC
TCGA-12-0826	TACC3	61DG9AAXX100217:4:114:5849:21081	R	1778470	1778602	33	1	0	TGCCATCCCCTAACACACACAGGGGGCTAAGGGCAGGG*** GAGGGGCATAGGCGGGGGCAGCCAGGCCAGAGC
TCGA-12-0826	TACC3	42MJN4AAXX00606:3:5:7078:8810#0	R	1778458	1778602	37	3	0	GAGGGGCATAGGCGGGGGGCACCAGGCAGAGGCAGC CACACACAGGGGGCTAAGGGCAGGG***GAGGGGCATAGCCG
TCGA-12-0826	TACC3	61DG9AAXX100217:3:56:4963:15876	R	1778454	1778602	52	5	0	GGGGGGACCAGGCCCGAGCCAGCACTCACGTCGGGGGG GACGTCCACCGACGTGAGTGCTGGCTCTGCCCTGGTGCCA
TCGA-12-0826	FGFR3	42MJN4AAXX00606:3:5:23:1551:1150#0	F	1778447	1778602	56	0	0	CCCGCCTATGCCCCTC***CCCTGCCCTTAGCCCCCCTG CTTACCCGTGACGTCCACCGACGTGAGTGCTGGCTCTGCCCT
TCGA-12-0826	FGFR3	61DG9AAXX100217:4:23:17613:20686	F	1778439	1778602	64	0	0	GGTGCCACCCGCCTATGCCCCTC***CCCTGCCCTTAG TGTCCTTACCGTGACGTCCACCGACGTGAGTGCTGGCTCTG
TCGA-12-0826	FGFR3	61DG9AAXX100217:4:23:14379:6949	F	1778435	1778602	68	0	0	GCCTGGTGCCACCCGCCTATGCCCCTC***CCCTGCCC
TCGA-19-5958	TACC3	D01RDACXX110628:6:1102:11167:161682	R	1778632	1778639	7	0	0	CCGGGGTGGGAGTGTGCCGGGTGACCCGCGGGTGTGGAGTGT GCAGGTGACCTCCCTGGCCCTTAGCCCCCCT***GCACTCT
TCGA-19-5958	TACC3	C01REACXX110628:2:2104:6309:98382	R	1778617	1778639	23	0	0	CCGGGTGACCGCGGGAGGGAGTGTGTGCAGGGGACCTCCCTG GCCCCTTAGCCCCCT***GCACTCTGTGGGGGCAGGATGGGC
TCGA-19-5958	TACC3	D01PPACXX110626:7:2103:12434:61998	R	1778603	1778639	39	0	0	TGAGTGTGCAGGTGACCTCCCTGGCCCTTAGCCCCCT***GC ACTCTGGGTGGGCAGGATGGCCGGGGACGGCAGGGGTGA
TCGA-27-1835	TACC3	B03UCABXX110625:8:1106:6252:48754	R	1778598	1778635	10	1	0	GGGGAGGCTGCTGGTGGGCAGCTGACTGCGGGGACACTGG GAGGAAGCCTGGACCCTGAGCGAAGT***TCGCCCAGCC
TCGA-27-1835	FGFR3	C00HWABXX110625:4:1201:20980:80977	F	1778595	1778635	29	0	0	ACAGCCTGGGCACAGAGGTTGCTGTGCGA***AGGTCGCTGA GGGTCCAGGCTTCCACCCACTTCCCCGCACGTCAGCT
TCGA-27-1835	TACC3	B03UC4BXX110625:5:1106:14043:83287	R	1778594	1778635	32	0	0	TGACTGCGGGGACACTTGGGTGGAAGCCTGGACCCTCAGCGG ACCT***TCGGCAGCCCACCTCTGTGCCCAGGCTGTGCC
TCGA-27-1835	TACC3	B097UABXX110405:4:2204:10446:98453	R	1778588	1778635	38	0	0	CGGGGCACTGGGTGGAAGCCTGGACCCTCAGCGACCT***T GGCAGAGCCACCTCTGTGCCCAGGCTGTGCCCCAGAA
TCGA-27-1835	TACC3	B097UABXX110405:4:2201:20858:44401	R	1778587	1778635	39	2	0	GGACTGGGTGGAAGCCTGGACCCTCAGCGACCT***TC GCACAGCCACCTCTGTGCCCAGGCTGTGCCCCAGAAG
TCGA-27-1835	TACC3	B097UABXX110405:2:2104:15656:71022	R	1778585	1778635	41	0	0	GGACACTGGGTGGAAGCCTGGACCCTCAGCGACCT***TCGC ACAGCCACCTCTGTGCCCAGGCTGTGCCCCAGAAGGC
TCGA-27-1835	TACC3	C00HWABXX110625:2:1102:20934:42427	R	1778543	1778635	53	3	0	GAAGCCTGGACCCTCAGCGACCT***TCGCACAGCCACCTCT GTGCCCCGGCTGTGCCCCAGCGGCGCCGCCCCACACC
TCGA-27-1835	TACC3	B09V2ABXX110405:8:1206:4774:81634	R	1778537	1778635	58	0	0	GAAGCCTGGACCCTCAGCGACCT***TCGCACAGCCACCTCT TGGACCCTCAGCGACCT***TCGCACAGCCACCTCTGTGCCCC
TCGA-27-1835	TACC3	C00HWABXX110625:7:2107:12251:167358	R	1778530	1778635	66	0	0	AGCCCTGTGCCCCAGAAGGCCCCGCCCCACACCTCACACACT TCAGCGACCT***TCGCACAGCCACCTCTGTGCCCCAGGCTGT
TCGA-27-1835	TACC3	B097UABXX110405:2:2104:15656:71022	R	1778523	1778635	73	0	0	GCT***TCGCACAGCCACCTCTGTGCCCCAGGCTGTGCCCCAG AAGCCCCGCCCCACACCTCACACACTTCAGCGACCTCGCA

Table 5 (Cont.)

sample	gene split2	read1 ID	direction split	hg18 start split2	hg18 stop split2	length2	mismatch2	gap2	seq mate
TCGA-06-5390	FGFR3	D03U9ACXX:110625:6:2205:12623:198252	R	1708787	1708854	68	4	0	GACGGTCCACCGACGTGAGTGCTTGGCTCTGGCCTGGTGCCA
TCGA-06-5390	TACC3	C01PRACXX:110625:6:1304:10052:66371	F	1708787	1708860	74	0	0	CCCGCCTATGCCCCTCCCCCTGCCGTCCCCGGCCAT
TCGA-06-5390	TACC3	C01PRACXX:110625:6:1304:10052:66371	F	1708787	1708860	74	0	0	CAAGAGCGGACTCAAGGACTTACAGGAATGTCCAGTGCTCCC
TCGA-06-5390	TACC3	C01PRACXX:110625:6:1108:21992	F	1708787	1708860	74	0	0	AAGAAATCGAACTCCACAAGCTTGGCTTCCCGCGG
TCGA-06-5390	TACC3	D01PRACXX:110625:6:2304:13007:106822	F	1708787	1708860	74	0	0	CAAGAGCGGACTCAAGGACTTACAGGAATGTCCAGTGCTCCC
TCGA-06-5390	TACC3	C01PRACXX:110625:6:2108:11999:91559	F	1708787	1708852	72	0	0	ATAGGCCCTTAAAACAACTCGTTCCCTCAGACCACACAGAAG
TCGA-06-5390	TACC3	C01PRACXX:110625:6:1306:1448:69211	F	1708787	1708856	70	0	0	ACAGTTCAAGAGCGGACTCAAGGACTTACAGGAAT
TCGA-06-5390	FGFR3	D03U9ACXX:110625:6:2205:12623:198252	R	1708787	1708854	68	4	0	TCAAGAGCGGACTCAAGGACTTACAGGAATGTCCAGTGCTCCC
TCGA-06-5390	TACC3	C01PRACXX:110625:6:2103:6915:17943	F	1708787	1708859	68	0	0	CAAGAAATCGAACTCCACAAGCTTGGCTTCCCGCGG
TCGA-06-5390	TACC3	C01PRACXX:110625:6:2204:8722:191886	F	1708787	1708852	66	0	0	ACCACACACAAGACAGTTCAAGAGCGGACTCAAGGACTTACA
TCGA-06-5390	TACC3	C01PRACXX:110625:6:1308:2911:26690	F	1708787	1708851	65	1	0	CGGCGAGCGAACAGCACAGCGGCCCCATGCCCCCAAC
TCGA-06-5390	TACC3	C01PRACXX:110625:6:2207:4585:84017	F	1708787	1708849	63	0	0	CGTTCCCTCAGACCACACACAAGACAGTTCAAGAGCGGACTC
TCGA-06-5390	FGFR3	C01PRACXX:110625:6:2205:11825:39734	R	1708787	1708841	55	0	0	AAGGACTTACAGGAATGTCCAGTGCTCCCAAGAGA
TCGA-06-5390	FGFR3	D01PRACXX:110625:6:1108:12159:179409	R	1708787	1708834	48	0	0	GCACCTGCCCCCAGGTACATGATCATGCCGGAGTTCCTGGCAT
TCGA-06-5390	FGFR3	D02U9ACXX:110625:4:2202:12501:80369	R	1708787	1708834	48	0	0	GCCCCGCCCCTCCCAGAGGCCCACCTTCAAGCAGCT
TCGA-06-5390	TACC3	C01RDACXX:110625:6:1305:2044:13239	F	1708787	1708813	27	0	0	CTGGCATGCCCCGCCCCTCCCAGAGGCCCACCTTTAAGCAGC
TCGA-06-5390	FGFR3	D03U9ACXX:110625:6:2205:13623:198552	R	1708787	1708840	54	4	0	TGGAGTTCGATTTCTTGGTAGCACTGGACATTCCTG
TCGA-06-5390	FGFR3	C01PRACXX:110625:6:2205:11825:39734	R	1708787	1708838	52	0	0	TCCCCCTGCCCGTCCCCGCCCATCCCTGAGGACGTCCCCCC
TCGA-06-5390	TACC3	D02U9ACXX:110625:7:2103:4492:173560	F	1708787	1708804	18	0	0	GAAGCCAAGCTTGTGGAGTTCGATTTCTTGGGAGCA
TCGA-06-5390	FGFR3	C01PRACXX:110625:6:2103:6915:17943	R	1708787	1708792	6	0	0	AGACCACACAGAAGACAGTTCAAGAGCGGACTCAAGGACTTA
TCGA-12-0826	FGFR3	C01C59AAXX:100217:4:1103:15123:61132	R	1707186	1707253	68	2	1	CAGGAATGTCCAGTGCTCCTCCAAGAGGCCCACCTTCAAGCAGCT
TCGA-12-0826	FGFR3	C01C59AAXX:100217:5:1107:10875:19649	R	1707186	1707252	67	2	1	GCCCCGCCCCTCCCAGAGGCCCACCTTCAAGCAGCT
TCGA-12-0826	TACC3	C01C59AAXX:100217:5:1106:19835:11296	F	1707186	1707252	67	2	1	GGTGGAGGACGTGGACCGTGTCCTTACCGTGACGTC
TCGA-12-0826	TACC3	C1C59AAXX:100217:5:1102:13123:10537	F	1707186	1707248	63	2	1	CGGCGCACATACCTGCTGGTGTCGGTGCCCACCGGCACTG
TCGA-12-0826	TACC3	42MJNAAXX:090913:5:23:1681:11150#0	F	1707186	1707236	51	2	1	GTCTACCAGGACTGTCCCTCAGGAGGGGTCAAACT
TCGA-12-0826	TACC3	C1C59AAXX:100217:5:1202:13123:10537	F	1707186	1707237	52	0	1	ATACCTGCTGGTCTCGGTGGCCACGGGCACTGGTCTACCAG
TCGA-12-0826	TACC3	42MJNAAXX:090913:5:23:1681:11150#0	F	1707186	1707236	51	2	1	GACTGTCCCTCAGGAGGGGTCAAACTTGAGGGTAT
TCGA-12-0826	TACC3	C1C59AAXX:100217:5:2101:17512:20592	F	1707186	1707237	52	0	1	AGGTATAAGGACTGCTTCCTCAAGGCCGACTCCTTAAACTGG
TCGA-12-0826	TACC3	5IC59AAXX:100217:4:2104:4270:6949	F	1707186	1707198	13	0	1	GGACAAGAGGGCAAGTGATGAGGTCTGACTGCCA
TCGA-12-0826	FGFR3	C01C59AAXX:100217:5:1107:11157:101462	R	1707186	1707270	85	2	1	GGAGGACGTGGACCGTGTCCTTACCGTGACGTCCACCGACG
TCGA-12-0826	FGFR3	C01REACXX:110625:2:2104:6009:96382	R	1707186	1707254	69	2	1	TGAGTGCTGGCTCTGGCCTGGTGCCACCCGGCGTAT
TCGA-12-0826	FGFR3	D01PRACXX:110625:7:2105:12434:81283	R	1707186	1707236	51	1	1	GGAGGACGTGGACCGTGTCCTTACCGTGACGTCCACCGACG
TCGA-19-5958	FGFR3	C01REACXX:110625:2:2104:6009:96382	R	1707202	1707254	53	2	0	TGAGTGCTGGCTCTGGCCTGGTGCCACCCGGCGTAT
TCGA-19-5958	FGFR3	D01PRACXX:110625:7:2105:12434:81283	R	1707202	1707236	35	1	0	GGAGGACGTGGACCGTGTCCTTACCGTGACGTCCACCGACG
TCGA-27-1835	FGFR3	B08UGABXX:110625:6:1102:9285:45754	R	1708367	1708462	96	2	0	TTCAAGCAGCTGGTGGAGGAGCCTGGACCGTGTCCTT
TCGA-27-1835	TACC3	C00HWABXX:110625:4:1201:20990:90677	F	1708367	1708449	47	0	0	GCCTGCACTGGGTCCTCAGGGGTGGGGGTCCCTCCGGGGCT
TCGA-27-1835	FGFR3	B08UGABXX:110625:5:1105:14043:85287	R	1708367	1708440	44	0	0	GGGCTGGGGGAGGGACTGGCAGGCCTGCAGGGGGGGTT
TCGA-27-1835	FGFR3	B097UABXX:110406:4:2204:12445:68453	R	1708367	1708404	36	0	0	TCACGGCAGCAAGAACACACTCACTGCTGCAAGTGCCACCA
TCGA-27-1835	FGFR3	B097UABXX:110406:4:2201:20652:94401	R	1708367	1708403	35	0	0	GAGGCCAACGCCATGCCCAGGCCGGAGAGTCCCCGG
TCGA-27-1835	FGFR3	B097UABXX:110406:2:2104:15668:71022	R	1708367	1708431	63	0	0	TACATGATCATGCCGGGAGGGCTGGCATGCCGCGCCTCCCA
TCGA-27-1835	FGFR3	C00HWABXX:110625:6:2102:20394:42457	R	1708367	1708419	53	0	0	GAGGCCCACCTTCAAGCAGCTGGTGGAGGAGCCTGGACCGTGTCCTCCA
TCGA-27-1835	FGFR3	B09V2ABXX:110406:6:1203:18187:141962	R	1708367	1708419	53	0	0	GGTGCCCCCCCAGAGTCCTGAGCGTGTGGGCCGGGCCTTCTG
TCGA-27-1835	FGFR3	B06V2ABXX:110406:8:1205:4774:81604	R	1708367	1708419	53	0	0	GCAGGTACATGATCATGCCGGGAGTGCCGGCATTTCGGGACCC
TCGA-27-1835	FGFR3	C00HWABXX:110625:2:1107:16163:23614	R	1708367	1708411	45	0	0	TTCCCTCGGGCCACCCTGTTCCGGTTGTTGTTGTGGC
TCGA-27-1835	FGFR3	C00HWABXX:110625:7:2107:12251:167363	R	1708367	1708408	42	0	0	GCAGGTACATGATCATGCCGGGAGTGCCGGCATTTCGGGACCC
TCGA-27-1835	FGFR3	B097UABXX:110406:2:2104:15938:71022	R	1708367	1708398	32	0	0	TCCCCGAGGACCACCTTCCAGCAGCCCGGCGAGGGC
									CCCGAATAAGGTGGGAAGCCGCCGGGCTGACTCCTGAGCCG
									CCCTGACCGCACGTACATGAGCATGCCGGGAGTGCCG
									CGTGTCCTTACCGTGACGTCCACCGACGTGAGTGCTTGGCTC
									TTGCCTTGCTGCCACCCGCCGTATGCCCCTCCCCCTG
									ACATGATCATGCCGGGAGTGCCTGCATGCCGCGCCTCCCCAG
									AGGCCCACCTTCAAGCAGCTGGTGGAGGAGCCTGGA
									GCCTTCTGGGGGCACAGCCTGGGCACACAGGTCGCTGTGCG
									AAGGTCGCTGAGGGTCCAGGCTTCCACCCAGTGTCC

[00265] The FGFR3 and TACC3 genes are located 48-Kb apart on human chromosome 4p16. The other members of the FGFR and TACC families retain the close physical association of FGFR3 and TACC3, with FGFR1 and TACC1 paired on chromosome 8p11 and FGFR2 and TACC2 paired on chromosome 10q26. Without being bound by theory, the ancestral FGFR and TACC genes were physically linked and that this tandem gene cluster

was duplicated at least twice to generate the FGFR1-TACC1, FGFR2-TACC2 and FGFR3-TACC3 pairs that mark mammalian evolution (Still et al., 1999). The highly conserved TK domains among FGFR genes and TACC domains among TACC genes together with their invariable fusion in the FGFR3-TACC3 rearrangements prompted to ask whether other intra-chromosomal FGFR-TACC fusion combinations exist in human GBM.

[00266] cDNA from a panel of 88 primary GBM were screened using pairs of upstream PCR primers that bind the amino-terminal coding region of the TK domains of FGFR1, FGFR2 and FGFR3 and downstream primers that bind to the carboxy-terminal coding region of the TACC domains of TACC1, TACC2 and TACC3 genes, respectively. The screening resulted in the identification of intrachromosomal FGFR-TACC fusions in two additional cases (one harboring FGFR1-TACC1 and one FGFR3-TACC3), corresponding to three of 97 total GBM (3.1%), including the GBM-1123 case. The FGFR1-TACC1 fusion breakpoint in GBM-51 joined in-frame exon 17 of FGFR1 to exon 7 of TACC1, resulting in a novel protein in which the TK domain of FGFR1 is fused upstream of the TACC domain of TACC1 (**Figure 2F**). The same structure was conserved again in GBM-22 in which exon 16 of FGFR3 is joined in-frame to exon 10 of TACC3 (**Figure 2G**). None of the tumors harboring FGFR-TACC fusions had mutations in IDH1 or IDH2 genes, thus indicating that FGFR-TACC-positive GBM mark an independent subgroup of patients from those carrying IDH mutations (Table 6) (Yan et al., 2009). The constant linkage of the FGFR-TK to the TACC domain created in each of the seven GBM harboring FGFR-TACC rearrangements suggests that FGFR-TACC fusion proteins may generate important functional consequences for oncogenesis in the brain.

[00267] **Table 6.**

Samples	Type	Time	Status	Age at initial pathologic diagnosis	IDH1-2 status (Sanger)	IDH1-2 status (exome)
TCGA-12-0826	FGFR3-TACC3	845	DECEASED	38	WT	WT
TCGA-27-1835	FGFR3-TACC3	648	DECEASED	53	NA	WT
TCGA-19-5958	FGFR3-TACC3	164	LIVING	56	NA	WT
TCGA-06-6390	FGFR3-TACC3	163	DECEASED	58	WT	WT
GBM-22	FGFR3-TACC3	396	DECEASED	60	WT	NA
GBM-1123	FGFR3-TACC3	NA	DECEASED	62	WT	NA
GBM-51	FGFR1-TACC1	NA	NA	NA	WT	NA

Time = Survival (days after diagnosis)
Sanger = analysis done by Sanger sequencing of genomic DNA
Exome = ainalysis done by the SAVI (Statistical Algorithm for Variant Identification), an algorithm developed to detect point mutation in cancer (BRAF Mutations in Hairy-Cell Leukemia, Tiacci E et al. The New England Journal of Medicine 2011 Jun 16;364(24):2305-15)
NA = Not Available
WT = Wild type sequence for R132 and R172 of IDH1 and IDH2, respectively

[00268] ***Transforming activity of FGFR-TACC fusions.*** To test the functional importance of the FGFR-TACC fusions in GBM, the FGFR3-TACC3 cDNA was cloned from GSC-1123 and recombinant lentiviruses were prepared expressing FGFR3-TACC3, FGFR1-TACC1, a kinase-dead FGFR3-TACC3 protein (FGFR3-TACC3-K508M), *wild type* FGFR3 and *wild type* TACC3. Transduction of Rat1A fibroblasts and *Ink4A;Arf-/-* astrocytes with the FGFR3-TACC3 lentivirus resulted in the expression of the fusion protein at levels comparable to those present in GSC-1123 (**Figure 11**). Having reconstituted in non-transformed cells the endogenous level of the FGFR-TACC protein that accumulates in GBM cells, it was determined whether it was sufficient to initiate oncogenic transformation *in vitro* and *in vivo*. Rat1A cells expressing FGFR3-TACC3 and FGFR1-TACC1 but not those expressing FGFR3-TACC3-K508M, FGFR3, TACC3 or the empty lentivirus acquired the ability to grow in anchorage-independent conditions in soft agar (**Figure 3A**). Transduction of the same lentiviruses in primary *Ink4A;Arf-/-* astrocytes followed by subcutaneous injection into immunodeficient mice revealed that only astrocytes expressing FGFR3-TACC3 and FGFR1-TACC1 formed tumors. The tumors emerged in 100% of the mice injected with astrocytes expressing the fusion proteins and were glioma-like lesions with strong positivity for Ki67, phospho-histone H3, nestin, GFAP and Olig2 (**Figure 3B**).

[00269] Next, it was determined whether the FGFR3-TACC3 fusion protein is oncogenic when transduced to a small number of cells directly into the brain of immunocompetent animals. A recently described mouse glioma model was used in which brain tumors are initiated by lentiviral transduction of oncogenes and inactivation of p53 in the mouse brain (Marumoto et al., 2009). To target adult NSCs, the adult mouse hippocampus was stereotactically transduced with purified lentivirus expressing the FGFR3-TACC3 protein and shRNA against p53 (pTomo-FGFR3-TACC3-shp53). Seven of eight mice (87.5%) transduced with FGFR3-TACC3 succumbed from malignant brain tumors within 240 days (**Figure 3C**). None of the mice transduced with a lentivirus expressing the most frequent gain-of-function mutation in GBM (the constitutively active EGFRvIII, pTomo-EGFRvIII-shp53) or the pTomo-shp53 control lentivirus died or developed clinical signs of brain tumors (**Figure 3C**). The FGFR3-TACC3 tumors were high-grade glioma with strong propensity to invade the normal brain and stained positive for the glioma stem cell markers nestin and Olig2 and the glial marker GFAP. They were also highly positive for Ki67 and phospho-histone H3, thus displaying rapid tumor growth (**Figure 3D**). The expression of FGFR3-TACC3 in the xenograft and intracranial tumor models was comparable to the expression of

the endogenous protein in the human GSCs and tumor (**Figure 11D, 11E and 11F**).

[00270] These data show that FGFR-TACC fusion proteins possess transforming activity in two independent cellular models and this activity is not the result of the overexpression of individual FGFR and TACC genes. They also show that direct transduction of the FGFR3-TACC3 protein to the adult mouse brain leads to efficient development of malignant glioma.

[00271] ***The FGFR-TACC fusions interfere with mitotic progression and induce chromosome missegregation and aneuploidy.*** To elucidate the mechanism by which the FGFR-TACC fusion drives oncogenesis, it was explored whether it activates downstream FGFR signaling. FGFR3-TACC3 failed to hyperactivate the canonical signaling events downstream of FGFR (pERK and pAKT) in the presence or absence of the ligands FGF-1, FGF-2 or FGF-8 (Wesche et al., 2011) (**Figures 12A, 12B and 12C**). However, FGFR3-TACC3 displayed constitutive phosphorylation of its TK domain and the adaptor protein FRS2, both of which were abolished by the specific inhibitor of FGFR-associated TK activity PD173074 (Mohammadi et al., 1998) or the K508M mutation (**Figure 4A**). Thus, FGFR3-TACC3 gains constitutive kinase activity that is essential for oncogenic transformation but the downstream signaling of this aberrant activity is distinct from the canonical signaling events downstream to FGFR. By driving the localization of the fusion protein, the TACC domain can create entirely novel TK-dependent functions. The TACC domain is essential for the localization of TACC proteins to the mitotic spindle (Hood and Royle, 2011; Peset and Vernos, 2008). Confocal imaging showed that FGFR3-TACC3 designed an arc-shaped structure bending over and encasing the metaphase spindle poles, frequently displaying asymmetry towards one of the two poles and relocated to the midbody as cells progressed into the late stages of mitosis (telophase and cytokinesis) (**Figure 4B and 12D**). Conversely, the localization of TACC3 was restricted to spindle microtubules and did not relocalize to the midbody (**Figure 12E**). Wild type FGFR3 lacked discrete localization patterns in mitosis (**Figure 12F**).

[00272] The mitotic localization of FGFR3-TACC3 indicates that it may impact the fidelity of mitosis and perturb the accurate delivery of the diploid chromosomal content to daughter cells, thus generating aneuploidy. Mitotic progression of individual cells was examined in vector-transduced and FGFR3-TACC3 expressing cells co-expressing histone H2B-GFP by time-lapse microscopy. The average time from nuclear envelope breakdown to anaphase onset was increased in cells expressing FGFR3-TACC3 in comparison with control

cells. The mitotic delay was further exacerbated by difficulties in completing cytokinesis (**Figures 4C and 4D**).

[00273] Next, it was determined whether the expressions of the FGFR-TACC fusion proteins induce defects of chromosomal segregation. Quantitative analyses of mitoses revealed that cells expressing FGFR3-TACC3 or FGFR1-TACC1 exhibit a three to five fold increase of chromosomal segregation errors than control cells. The most frequent mitotic aberrations triggered by the fusion proteins were misaligned chromosomes during metaphase, lagging chromosomes at anaphase and chromosome bridges that impaired cytokinesis and generated micronuclei in the daughter cells (**Figures 4E, 4F and 13A**). Aberrations at the metaphase-anaphase transition frequently lead to the inability of mitotic cells to maintain a metaphase arrest after treatment with a spindle poison. Over 18% of cells expressing FGFR3-TACC3 displayed prematurely separated sister chromatids in contrast with less than 3% in control, FGFR3 or TACC3-expressing cells (**Figures 13B and 13C**). Accordingly, cells expressing the fusion protein were unable to efficiently arrest in metaphase after nocodazole treatment (**Figure 13D**).

[00274] The above findings indicate that expression of the FGFR3-TACC3 fusion protein may spark aneuploidy. Karyotype analysis revealed that FGFR3-TACC3 increased over 2.5 fold the percent of aneuploidy and led to the accumulation of cells with broad distribution of chromosome counts in comparison with cells transduced with empty vector, FGFR3 or TACC3 (**Figure 5A**). Accordingly, GSC-1123 contained aneuploid modal number of chromosomes (49) and manifested a broad distribution of chromosome counts characterized by 60% of metaphase spreads that deviate from the mode (Table 7).

[00275] Table 7: Chromosome analysis by SKY of 20 cells from the GSC-1123 culture

Cell #	Chr #	+1	+2	+3	t(3;14)	+4	(-4)	del(4)	+5	+6	+7	del(7)	+8	-9	+9	-10	+10	+11
1	97	2	2	2		2		1	2	2	4		2		2			2
2	51										1					1		
3	49										1					1		
4	50										1					1		
5	49										1					1		
6	86	2	2	2		2			2		3		1		1			2
7	95	2	2	2		2			1	2	4		2		2			2
8	98	2	3	2	1	2			1	2	4		3		1			2
9	86	2	1	1	1	1			1	1	6		2		3		1	1
10	44			1	1		1				1	1		1		1		
11	49										1	1				1		
12	49										1					1		
13	98	2	2	2	2				2	2	4		2		2			2
14	49										1					1		
15	48										1					1		
16	51										1					1		
17	49										1					1		
18	50										1	1				1		
19	49										1					1		
20	49										1					1		

Cell #	Chr #	+12	+13	del(13)	-14	+14	+15	-16	+16	+17	+18	+19	+20	-21	+21	-22	+22	+X
1	97	2	2	2		2	2		2	2	3	4	4		2		2	2
2	51		2	2							1	1	1					
3	49			1							1	1	1					
4	50			1							1	1	1					1
5	49			1							1	1	1					
6	86	2	2	2		2	2			2	4	3	3		1			2
7	95	2	2	2		2	2		2	2	4	2	4		2		2	2
8	98	2	2	2		2	2		2	2	3	4	4		3		2	2
9	86	2	2	2					2	1	4	3	1		2		1	2
10	44				1			1			1		1			1		
11	49										1	1	1					
12	49			1							1	1	1					
13	98	2	2			2	2		2	2	4	4	4		2	2		2
14	49			1							1	1	1					
15	48										1	1	1	1				
16	51			1		1					1						2	1
17	49		1								1	1	1					
18	50		1	1							1	1	1					
19	49										1	1	1					
20	49			1							1	1	1					

[00276] Next, it was determined whether aneuploidy is a direct consequence of FGFR3-TACC3 expression and is induced in human diploid neural cells. Primary human astrocytes analyzed six days after transduction with the FGFR3-TACC3 lentivirus exhibited a 5-fold increase of the rate of aneuploidy and a significantly wider distribution of chromosome counts (**Figures 5B, 5C and 5D**). Consistent that aneuploidy is detrimental to cellular fitness, acute expression of FGFR3-TACC3 compromised the proliferation capacity of human astrocytes. However, continuous culture of FGFR3-TACC3-expressing human astrocytes led to progressive gain of proliferative capacity that overrode that of control cells (**Figure 14A, 14B**). Thus, the acute expression of FGFR3-TACC3 in primary normal human cells from the central nervous system causes CIN and aneuploidy with an acute fitness cost manifested by

slower proliferation.

[00277] It was also determined whether the CIN and aneuploidy caused by FGFR3-TACC3 requires the TK activity of FGFR3 and can be corrected. Treatment with PD173074 rescued the aneuploidy caused by FGFR3-TACC3 by over 80%, restored the narrow distribution of chromosome counts typical of control cells and largely corrected the cohesion defect (**Figures 6A, 6B and 6C**). Together, these findings indicate that the CIN and aneuploidy caused by rearrangements of FGFR and TACC genes are reversible and suggest that specific FGFR kinase inhibition may be a valuable therapeutic strategy in tumor cells expressing FGFR-TACC fusion proteins.

[00278] ***FGFR-TACC fusion proteins are new therapeutic targets in GBM.*** Driver genetic alterations trigger a state of oncogene addiction in the cancer cells harboring them that can be exploited therapeutically. To ask whether FGFR-TACC fusions confer addiction to FGFR-TK activity, cell growth was analyzed in the presence of PD173074, AZD4547 or BGJ398, the latter being two highly specific inhibitors of FGFR-TK under clinical investigation (Gavine et al., 2012; Guagnano et al., 2011). Each of the three drugs inhibited growth of cells expressing FGFR3-TACC3 and FGFR1-TACC1 at concentrations <10 nM whereas they were ineffective at concentrations as high as 1 μM in cells transduced with vector, FGFR3, TACC3 and the FGFR3-TACC3-K508M mutant (**Figures 7A, 14C** and **14D**). These findings underscore the elevated degree of specificity for FGFR kinase inhibition towards cells carrying the fusion protein. The growth of GSC-1123 cells, which naturally harbor the FGFR3-TACC3 translocation, was also abolished by nanomolar concentrations of FGFR-TK inhibitors (**Figure 7B**). Targeting of the fusion gene by FGFR3 shRNA inhibited the growth of cells ectopically expressing FGFR3-TACC3 and GSC-1123 proportionally to the silencing efficiency of FGFR3-TACC3 (**Figures 7C and 14E**).

[00279] Finally, it was determined whether treatment with PD173074 of mice bearing glioma xenografts of FGFR3-TACC3 transformed astrocytes inhibits tumor growth. Twelve days after injection of tumor cells, subcutaneous tumors were present in all animals. The mice were randomized in two cohorts and treated with PD173074 or vehicle. PD173074 elicited a potent growth inhibition of FGFR3-TACC3 glioma (**Figure 7D**). To confirm the efficacy of a clinically meaningful FGFR-TK inhibitor using a more anatomically relevant model, the AZD4547 FGFR inhibitor, a compound under clinical investigation (Gavine et al., 2012), was used against intracranial luciferase-expressing FGFR3-TACC3-driven glioma

xenografts. After an engraftment period, tumor-bearing animals were treated with either AZD4547 or vehicle. Oral administration of AZD4547 markedly prolonged survival (**Figure 7E**). Taken together, the data provide a strong rationale for a clinical trial based on FGFR inhibitors in GBM harboring FGFR-TACC rearrangements.

[00280] **DISCUSSION**

[00281] This work has established that recurrent, oncogenic and addicting gene fusions identify a subset of GBM patients. The functional characterization of FGFR-TACC fusions indicates that the constitutively active FGFR-TK and the TACC domain of the fusion protein are both essential for oncogenesis. The TACC-dependent mis-localization to mitotic cells of the FGFR kinase results in aberrant compartmentalization of a constitutively active TK to the mitotic spindle pole, thus providing a mechanistic explanation for the impaired mitotic fidelity, chromosome mis-segregation and aneuploidy instigated by the fusion protein.

[00282] Without being bound by theory, mutation of the genes that control chromosome segregation during mitosis can explain the high rate of CIN and aneuploidy, which is typical of most solid tumors including GBM (Gordon et al., 2012). A few examples of mutational inactivation of candidate genes have been reported in human cancer (Solomon et al., 2011; Thompson et al., 2010). However, gain-of-function mutations causally implicated in the control of mitotic fidelity have not been described. This clashes with the classical observation from cell fusion experiments that the underlying mechanisms that cause CIN behave as dominant traits, indicating that the CIN phenotype results from gain-of-function events rather than gene inactivation (Lengauer et al., 1997, 1998). The FGFR-TACC gene fusion is a novel mechanism for the initiation of CIN and provides a clue to the nature of dominant mutations responsible for aneuploidy in human cancer.

[00283] The rapid emergence of mitotic defects and aneuploid cell populations triggered by the fusion protein in normal human astrocytes, combined with the correction of aneuploidy after short inhibition of FGFR-TK activity indicate that aneuploidy is a key event in tumor induction by the FGFR-TACC gene fusions. Induction of aneuploidy *per se* is detrimental to cellular fitness (Sheltzer and Amon, 2011). Full-blown tumorigenesis requires cooperation between aneuploidy and genetic lesions that confer growth advantage and protect cells against the detrimental effects of aneuploidy (Coschi and Dick, 2012; Holland and Cleveland, 2009; Weaver and Cleveland, 2009). Therefore, the potent tumor-initiating

activity of FGFR-TACC fusions shows that the novel oncoproteins have growth-promoting signaling functions that complement the loss of mitotic fidelity with ensuing karyotypic alterations (Sheltzer and Amon, 2011).

[00284] Targeted therapies against common genetic alterations in GBM have not changed the dismal clinical outcome of the disease, most likely because they have systematically failed to eradicate the truly addicting oncoprotein activities of GBM. The highly specific anti-tumor effects and the correction of aneuploidy precipitated by FGFR-TK inhibition of FGFR-TACC-driven GBM provide a strong rationale for clinical trials based on FGFR inhibitors in patients harboring FGFR-TACC rearrangements. The computational gene fusion discovery pipeline reported here detected other GBM cases in which FGFR family genes are implicated in additional gene fusions beyond the FGFR-TACC rearrangements. Therefore, the frequency of 3.1% is likely to be an underestimate of the target GBM patient population that may benefit from FGFR-TK inhibition.

[00285] **EXPERIMENTAL PROCEDURES**

[00286] ***Cell culture and isolation and maintenance of GSCs.*** Rat1A, mouse astrocytes *Ink4A;Arf-/-*, and human astrocytes were cultured in DMEM supplemented with 10% FBS. Isolation and culture of GSCs was performed as described (Carro et al., 2010). For treatment *in vitro* with PD173074, AZD4547 or BJG398, cells infected with vector control, FGFR3, TACC3, FGFR-TACC fusions or FGFR3-TACC3-K508M were seeded in 96-well plates and treated with increasing concentrations of FGFR inhibitors. After 72-120 h, growth rate was measured using the 3-(4,5-dimethythiazol-2-yl)-2,5-diphenyl tetrazolium bromide (MTT) assay. Data were expressed as mean±SD. Proliferation rate in GSC-1123 infected with FGFR3 shRNA lentivirus was determined by plating dissociated gliomaspheres at $2x10^4$ cells/well in twelve-well plates 5 days after infection. The number of viable cells was determined by trypan blue exclusion in triplicate cultures obtained from triplicate independent infections. Cell number was scored every other day.

[00287] ***DNA, RNA preparation, genomic and real-time quantitative PCR (qRT-PCR).***

[00288] The validation of fusion transcripts was performed using both genomic and RT-PCR with forward and reverse primer combinations designed within the margins of the paired-end read sequences detected by RNA-seq. DNA, RNA preparation and qRT-PCR were performed as described (Carro et al., 2010; Zhao et al., 2008). To identify novel fusion

transcripts within the GBM cohort, PCR primers pairs were designed to bind upstream to the TK domain of the FGFR genes and inside or downstream the Coiled Coil domain of the TACC genes. Expressed fusion transcript variants were subjected to direct sequencing to confirm sequence and frame. Primer sequences are included below.

[00289] ***Subcutaneous xenografts and drug treatment.*** Rat1A or *Ink4A;Arf-/-* astrocytes ($5x10^5$) transduced with different lentiviral constructs were suspended in 150 µl of PBS, together with 30 µl of Matrigel (BD Biosciences), and injected subcutaneously in the flank of athymic nude (Nu/Nu) mice (Charles River Laboratories, Wilmington, MA). For experiments with FGFR inhibitors, mice carrying ~200-300 mm^3 subcutaneous tumors derived from *Ink4A;Arf-/-* astrocytes were randomized to receive 50 mg/kg PD173074 in 0.05 M lactate buffer (pH 5) or an equal volume of lactate buffer by oral gavage. Treatment was administered for three cycles consisting of four consecutive days followed by two days of rest. Tumor diameters were measured with caliper, and tumor volumes estimated using the formula: 0.5 x length x $width^2$. Data are expressed as mean±SE. Mice were sacrificed when tumors in the control group reached the maximal size allowed.

[00290] ***Orthotopic transplantation and drug treatment.*** *Ink4A;Arf-/-* astrocytes carrying a luciferase expressing vector were transduced with FGFR3-TACC3 lentivirus. $1x10^3$ cells in 2 µl of saline were injected in the caudate-putamen of 4-6 week old male athymic nude (Nu/Nu) mice using a stereotaxic frame (coordinates relative to bregma: 0.5 mm anterior; 1.1 mm lateral; 3.0 mm ventral) and a 26 gauge Hamilton syringe. Six days after injection, mice underwent bioluminescence imaging using a Xenogen CCD apparatus and were randomized to receive 50 mg/kg AZD4547 in 1% Tween 80 (treatment group) or DMSO in an equal volume of vehicle by oral gavage (control group). AZD4547 was administered daily for two cycles of 10 days with a two day interval. Mice were monitored daily and sacrificed when neurological symptoms appeared. Kaplan-Meier survival curve was generated using the DNA Statview software package (AbacusConcepts, Berkeley, CA). Log-rank analysis was performed on the Kaplan-Meier survival curve to determine statistical significance.

[00291] ***Intracranial injections of lentiviruses.*** Intracranial injection of FGFR3-TACC3-shp53, EGFRvIII-shp53 or shp53 pTomo lentiviruses was performed in 4 week-old C57/BL/6J mice in accordance with guidelines of IACUC Committee. Briefly, 1.8 µl of purified lentiviral particles in PBS ($1X10^9$/ml) were injected into the dentate gyrus using a stereotaxic frame (coordinates relative to bregma: 1.45 mm posterior; 1.65 mm lateral; 2.4

mm ventral) and a 26 gauge Hamilton syringe. Mice were monitored daily and sacrificed when neurological symptoms appeared. Mouse brain was analyzed histopathologically and by immunofluorescence staining.

[00292] ***Histology and immunostaining.*** Tissue preparation and immunohistochemistry on brain tumors and immunofluorescence staining were performed as previously described (Carro et al., 2010; Zhao et al., 2009; Zhao et al., 2008). Antibodies used in immunostaining and immunoblotting are listed below.

[00293] ***Cloning and Lentiviral production.*** Lentivirus preparation and infections were performed as described (Carro et al., 2010) and are detailed in Extended Experimental Procedures.

[00294] ***Karyotype analysis.*** Cultured cells were colcemid (20 ng/ml) treated for 90 minutes before harvesting for karyotopic analysis as detailed in Extended Experimental procedures. At least one hundred cells in metaphase were examined for chromosome count. PMSCS was scored in cells where a majority of the sister chromosomes were no longer associated. Two-tailed unpaired t-tests with Welch's correction were performed for comparison of means analysis.

[00295] ***Immunofluorescence and live-cell microscopy.*** Immunofluorescence microscopy was performed on cells fixed with 4% PFA in PHEM (60 mM Pipes, 27 mM Hepes, 10 mM EGTA, 4 mM $MgSO_4$, pH 7.0). Cells were permeabilized using 1% Triton X-100. Mitotic spindles were visualized by anti-α-tubulin antibody (Sigma). Secondary antibodies conjugated to Alexa Fluor-488/-594 (Molecular Probes) were used. All staining with multiple antibodies were performed in a sequential manner. DNA was stained by DAPI (Sigma). Fluorescence microscopy was performed on a Nikon A1R MP microscope.

[00296] ***Identification of gene fusions from whole transcriptome (RNA-seq) and exome sequencing.*** RNA-Sequencing was performed from total RNA extracted from GSC cultures isolated from nine GBM patients using Illumina HiSeq 2000, producing roughly 60.3 million paired reads per sample. Using the global alignment software Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2009) with modified Mott's trimming, an initial seed length of 32, maximum edit distance of 2 and a maximum gap number of 1, on average 43.1 million reads were mapped properly to the RefSeq transcriptome and, of the remaining, 8.6 million were mapped to the hg19 genome per sample. The remaining 14.3% of paired reads—including

those that failed to map to either transcriptome or genome with proper forward-reverse (F-R) orientation, within expected insert size, and with minimal soft clipping (unmapped portions at the ends of a read)— were considered to be appropriate for gene fusion analysis.

[00297] A novel computational pipeline was constructed called *TX-Fuse* that identifies two sources of evidence for the presence of a gene fusion: 1. *Split inserts*, in which each read of a mate pair maps entirely to one side of a breakpoint, and 2. Individual *split reads* that span a breakpoint. Split inserts are readily detected from BWA mapping. On the other hand, split reads demand precision alignment of smaller nucleotide stretches. To that end, the pipeline employs the local alignment package BLAST with word size of 20, identity cutoff of 95%, expectation cutoff of 10^{-4}, and soft filtering to map raw paired reads against the RefSeq transcriptome. From this procedure, a list of potential split reads were obtained that were filtered to ensure maintenance of coding frame in the predicted fusion transcript given the proper F-R orientation in the read pair. False positive candidates produced from paralogous gene pairs were also screened out using the Duplicated Genes Database and the EnsemblCompara GeneTrees (Vilella et al., 2009). Pseudogenes in the candidate list were annotated using the list from HUGO Gene Nomenclature Committee (HGNC) database (Seal et al., 2011) and given lower priority. For each remaining gene fusion candidate, a virtual reference was created based on the predicted fusion transcript and re-mapped all unmapped reads using BLAST with word size of 16, identity cutoff of 85%, query coverage greater than 85%, and expectation cutoff of 10^{-4} to obtain a final count of split reads and inserts. Moreover, sequencing depth per base of the virtual reference was calculated to corroborate that components of each gene participating in the gene fusion were highly expressed.

[00298] To establish the recurrence of the initial panel of gene fusion candidates, the gene fusion discovery pipeline was modified to produce *EXome-Fuse*, which probes for fusions within the available dataset of paired-read exome DNA sequencing of 84 matched GBM samples from TCGA. To increase sensitivity for gene fusion identification, reads unmapped by BWA were aligned to the gene pair participating in each fusion candidate using a BLAST word size of 24 for split inserts and 16 for split read and split insert discovery. Given that the breakpoint detected in DNA cannot directly indicate the resulting breakpoint in the transcribed RNA, no restriction was made on split insert orientation. For split reads, it was only required that the component of the split read mapped to the same gene as its mate maintained F-R directionality.

[00299] ***Co-outlier expression and CNV analysis from TCGA GBM samples.*** Tomlins et al. (Tomlins et al., 2005) reported that outlier gene expression from microarray datasets identifies candidate oncogenic gene fusions. Wang et al. (Wang et al., 2009) suggested a "breakpoint principle" for intragenic copy number aberrations in fusion partners. The two principles (outlier expression and intragenic CNV) were combined to identify candidate gene fusions in GBM samples from Atlas-TCGA. Genomic and expression data sets were downloaded from TCGA public data portal as available on December 1, 2011, where a description of TCGA data types, platforms, and analyses is also available (2008). Specific data sources were (according to Data Levels and Data Types) as follows: Expression data, "Level 2" normalized signals per probe set (Affymetrix HT_HG-U133A) of 84 samples; Copy number data, "Level 1" raw signals per probe (Affymetrix Genome-Wide Human SNP *Array* 6.0) of the 4 *FGFR3-TACC3* gene fusion positive samples (tumor and matched normal control).

[00300] The gene expression analysis was performed first using R^3. The median absolute deviation (MAD) was calculated and then a gene was labeled as an outlier according to the following formula: $Z_{i,j} = 0.6745(x_{i,j} - \text{mean}(x_i))/MAD_i > 3.5$ (Iglewicz and Hoaglin, 1993). Samples were identified as ECFS (expression candidate fusion sample) if both genes of interest (e. g. FGFR3 and TACC3) displayed outlier behavior (co-outliers). Next, ECFS were analyzed for CNV using pennCNV (Wang et al., 2007). Tumors samples were paired to their normal controls to obtain the log ratio values and the VEGA algorithm was used to obtain a more accurate segmentation (Morganella et al., 2010).

[00301] ***Karyotypic Analysis.*** The colcemid treated cells were trypsinized, centrifuged for 7 minutes at 200 x g, and the cell pellet re-suspended in warmed hypotonic solution and incubated at 37°C for 13 minutes. The swollen cells were then centrifuged and the pellet re-suspended in 8 ml of Carnoy's fixative (3:1 methanol:glacial acetic acid). The cell suspension was centrifuged and washed twice in Carnoy's fixative. After the last centrifugation, the cells were resuspended in 0.5 to 1 ml of freshly prepared fixative to produce an opalescent cell suspension. Drops of the final cell suspension were placed on clean slides and air-dried. Slides were stained with DAPI and metaphases were analyzed under a fluorescent microscope.

[00302] **Cloning and Lentiviral production.** Lentiviral expression vectors, pLOC-GFP (Open Biosystems) and pTomo-shp53, were used to clone FGFR3, TACC3, FGFR3-TACC3,

FGFR3-TACC3-K508M, and FGFR1-TACC1. pTomo-shp53 was a gift of Inder Verma and Dinorah Friedman-Morvinski (Salk Institute, San Diego). The FGFR3-TACC3-K508M mutant was generated using the Phusion Site Direct Mutagenesis kit (NEB, USA). MISSION shRNAs clones (pLKO.1 lentiviral expression vectors) against FGFR3 were purchased from Sigma. The hairpin sequences targeting the *FGFR3* gene are-

[00303] 5′-TGCGTCGTGGAGAACAAGTTT-3′ (#TRCN0000000372; Sh#2) (SEQ ID NO: 182);

[00304] 5′-GTTCCACTGCAAGGTGTACAG-3′ (#TRCN0000430673; Sh#3) (SEQ ID NO: 183);

[00305] 5′-GCACAACCTCGACTACTACAA-3′ (#TRCN0000000374; Sh#4) (SEQ ID NO: 184).

[00306] **Genomic and mRNA RT-PCR.** Total RNA was extracted from cells by using RNeasy Mini Kit (QIAGEN), following the manufacturer instructions. 500 ng of total RNA was retro-transcribed by using the Superscript III kit (Invitrogen), following the manufacturer instructions. The cDNAs obtained after the retro-transcription was used as templates for qPCR. The reaction was performed with a Roche480 thermal cycler, by using the Absolute Blue QPCR SYBR Green Mix from Thermo Scientific. The relative amount of specific mRNA was normalized to 18S. Results are presented as the mean ± SD of triplicate amplifications.

[00307] Primers used are:

hFGFR3-RT-FW1: 5'-GTAACCTGCGGGAGTTTCTG-3' (SEQ ID NO: 162);

hFGFR3-RT-REV1: 5'-ACACCAGGTCCTTGAAGGTG-3' (SEQ ID NO: 163);

hTACC3-RT-FW2: 5'-CCTGAGGGACAGTCCTGGTA-3' (SEQ ID NO: 164);

hTACC3-RT-REV2: 5'-AGTGCTCCCAAGAAATCGAA-3' (SEQ ID NO: 165);

hWRAP53-RT-FW1: 5'-AGAGGTGACCACCAATCAGC-3' (SEQ ID NO: 180);

hWRAP53-RT-REV1: 5'-CGTGTCCCACACAGAGACAG-3' (SEQ ID NO: 181).

[00308] Primers used for the screening of FGFR-TACC fusions are:

FGFR3-FW1: 5'-CGTGAAGATGCTGAAAGACGATG-3' (SEQ ID NO: 166);

TACC3-REV1: 5'- AAACGCTTGAAGAGGTCGGAG-3' (SEQ ID NO: 167);

FGFR1-FW1: 5'-ATGCTAGCAGGGGTCTCTGA-3' (SEQ ID NO: 168);

TACC1-REV1: 5'-CCCTTCCAGAACACCTTTCA-3' (SEQ ID NO: 169).

[00309] Primers used for genomic detection of FGFR3-TACC3 fusion in GBM-1123 and GSC-1123 are:

Genomic FGFR3-FW1: 5'-ATGATCATGCGGGAGTGC-3' (SEQ ID NO: 170);

genomicTACC3-REV1: 5'-GGGGGTCGAACTTGAGGTAT-3' (SEQ ID NO: 171).

[00310] Primers used to validate fusions detected by RNA-seq are:

POLR2A-FW1: 5'-CGCAGGCTTTTTGTAGTGAG-3' (SEQ ID NO: 172);

WRAP53-REV1: 5'-TGTAGGCGCGAAAGGAAG-3' (SEQ ID NO: 173);

PIGU-FW1: 5'-GAACTCATCCGGACCCCTAT-3' (SEQ ID NO: 174);

NCOA6-REV1: 5'-GCTTTCCCCATTGCACTTTA-3' (SEQ ID NO: 175);

ST8SIA4-FW1: 5'-GAGGAGAGAAGCACGTGGAG-3' (SEQ ID NO: 176);

PAM-REV1: 5'-GGCAGACGTGTGAGGTGTAA-3' (SEQ ID NO: 177);

CAPZB-FW: 5'-GTGATCAGCAGCTGGACTGT-3' (SEQ ID NO: 178);

UBR4-REV1: 5'-GAGCCTGGGCATGGATCT-5' (SEQ ID NO: 179).

[00311] **Confocal microscopy imaging.** For immunofluorescence of fixed cells, images were recorded with a Z-optical spacing of 0.25 μm using a Nikon A1R MP and a 60X1.3 oil objective and analyzed using ImageJ software (National Institute of Health). For live-cell analyses, Rat1A cells infected with pLNCX-H2B retrovirus and transduced with lentiviral vector or FGFR3-TACC3 fusion were seeded in glass bottom dishes in phenol red free DMEM and followed by time-lapse microscopy using the Nikon A1R MP biostation at 37°C and 5% CO_2/95% air. Images with a Z-optical spacing of 1 μm were recorded every 4 min for 8 h. Images of unchallenged mitosis from early prophase until cytokinesis were processed using ImageJ software (National Institute of Health). The time-point of nuclear envelope

breakdown (NEB) was defined as the first frame showing loss of smooth appearance of chromatin and anaphase was the first frame when chromosome movement towards the poles became apparent. Nuclear envelope reconstitution (NER) was defined as the first frame showing nuclei decondensation.

[00312] Box and whisker plots were calculated from image sequences from at least 50 recorded cells. Two-tailed unpaired t-tests with Welch's correction were performed for comparison of means analysis using StatView software (AbacusConcepts, Berkeley, CA).

[00313] **Immunofluorescence.** Antibodies and concentrations used in immunofluorescence staining are:

Anti-Ki67	Rabbit	1:1000	Vector Labs
Anti-pHH3	Rabbit	1:500	Millipore
Anti-FGFR3	Mouse	1:1000	Santa Cruz
Anti-Tacc3	Goat	1:1000	USBiological
Anti-a-tubulin	Mouse	1:1000	Sigma
Anti-Nestin	Mouse	1:1000	BD Pharmingen
Anti-Olig2	Rabbit	1:200	IBL
Anti-GFAP	Rabbit	1:200	Dako
Anti-ERK	Rabbit	1:1000	Cell Signaling
Anti-pERK	Rabbit	1:1000	Cell Signaling
AntiFRS	Rabbit	1:250	Santa Cruz
Anti-pFRS	Rabbit	1:1000	Cell Signaling
Anti-AKT	Rabbit	1:1000	Cell Signaling
Anti-pAKT473	Rabbit	1:1000	Cell Signaling

[00314] **REFERENCES**

Ablain, J., Nasr, R., Bazarbachi, A., and de The, H. (2011). The Drug-Induced Degradation of Oncoproteins: An Unexpected Achilles' Heel of Cancer Cells? Cancer Discov. *1*, 117-127.

Bass, A.J., Lawrence, M.S., Brace, L.E., Ramos, A.H., Drier, Y., Cibulskis, K., Sougnez, C., Voet, D., Saksena, G., Sivachenko, A., *et al.* (2011). Genomic sequencing of colorectal adenocarcinomas identifies a recurrent VTI1A-TCF7L2 fusion. Nat. Genet. *43*, 964-968.

Cahill, D.P., Kinzler, K.W., Vogelstein, B., and Lengauer, C. (1999). Genetic instability and darwinian selection in tumours. Trends Cell. Biol. *9*, M57-60.

Carro, M.S., Lim, W.K., Alvarez, M.J., Bollo, R.J., Zhao, X., Snyder, E.Y., Sulman, E.P., Anne, S.L., Doetsch, F., Colman, H., *et al.* (2010). The transcriptional network for mesenchymal transformation of brain tumours. Nature *463*, 318-325.

Coschi, C.H., and Dick, F.A. (2012). Chromosome instability and deregulated proliferation: an unavoidable duo. Cell. Mol. Life Sci. *69*, 2009-2024

Druker, B.J. (2009). Perspectives on the development of imatinib and the future of cancer research. Nat. Med. *15*, 1149-1152.

Furnari, F.B., Fenton, T., Bachoo, R.M., Mukasa, A., Stommel, J.M., Stegh, A., Hahn, W.C., Ligon, K.L., Louis, D.N., Brennan, C., *et al.* (2007). Malignant astrocytic glioma: genetics, biology, and paths to treatment. Genes Dev. *21*, 2683-2710.

Gavine, P.R., Mooney, L., Kilgour, E., Thomas, A.P., Al-Kadhimi, K., Beck, S., Rooney, C., Coleman, T., Baker, D., Mellor, M.J., *et al.* (2012). AZD4547: An Orally Bioavailable, Potent, and Selective Inhibitor of the Fibroblast Growth Factor Receptor Tyrosine Kinase Family. Cancer Res. *72*, 2045-2056.

Gerber, D.E., and Minna, J.D. (2010). ALK inhibition for non-small cell lung cancer: from discovery to therapy in record time. Cancer Cell *18*, 548-551.

Gordon, D.J., Resio, B., and Pellman, D. (2012). Causes and consequences of aneuploidy in cancer. Nature reviews Genet. *13*, 189-203.

Guagnano, V., Furet, P., Spanka, C., Bordas, V., Le Douget, M., Stamm, C., Brueggen, J., Jensen, M.R., Schnell, C., Schmid, H., *et al.* (2011). Discovery of 3-(2,6-dichloro-3,5-dimethoxy-phenyl)-1-{6-[4-(4-ethyl-piperazin-1-yl)-phenylamin o]-pyrimidin-4-yl}-1-methyl-urea (NVP-BGJ398), a potent and selective inhibitor of the fibroblast growth factor receptor family of receptor tyrosine kinase. J. Med. Chem. *54*, 7066-7083.

Holland, A.J., and Cleveland, D.W. (2009). Boveri revisited: chromosomal instability, aneuploidy and tumorigenesis. Nat. Rev. Mol. Cell. Biol. *10*, 478-487.

Hood, F.E., and Royle, S.J. (2011). Pulling it together: The mitotic function of TACC3. Bioarchitecture *1*, 105-109.

Lee, J., Kotliarova, S., Kotliarov, Y., Li, A., Su, Q., Donin, N.M., Pastorino, S., Purow, B.W., Christopher, N., Zhang, W., *et al.* (2006). Tumor stem cells derived from glioblastomas cultured in bFGF and EGF more closely mirror the phenotype and genotype of primary tumors than do serum-cultured cell lines. Cancer Cell *9*, 391-403.

Lengauer, C., Kinzler, K.W., and Vogelstein, B. (1997). Genetic instability in colorectal cancers. Nature *386*, 623-627.

Lengauer, C., Kinzler, K.W., and Vogelstein, B. (1998). Genetic instabilities in human cancers. Nature *396*, 643-649.

Lo, H.W. (2010). EGFR-targeted therapy in malignant glioma: novel aspects and mechanisms of drug resistance. Curr. Mol. Pharmacol. *3*, 37-52.

Marumoto, T., Tashiro, A., Friedmann-Morvinski, D., Scadeng, M., Soda, Y., Gage, F.H., and Verma, I.M. (2009). Development of a novel mouse glioma model using lentiviral vectors. Nat. Med. *15*, 110-116.

Mitelman, F., Johansson, B., and Mertens, F. (2007). The impact of translocations and gene fusions on cancer causation. Nat. Rev. Cancer *7*, 233-245.

Mohammadi, M., Froum, S., Hamby, J.M., Schroeder, M.C., Panek, R.L., Lu, G.H., Eliseenkova, A.V., Green, D., Schlessinger, J., and Hubbard, S.R. (1998). Crystal structure of an angiogenesis inhibitor bound to the FGF receptor tyrosine kinase domain. EMBO J. *17*, 5896-5904.

Ohgaki, H., and Kleihues, P. (2005). Population-based studies on incidence, survival rates, and genetic alterations in astrocytic and oligodendroglial gliomas. J. Neuropathol. Exp. Neurol. *64*, 479-489.

Peset, I., and Vernos, I. (2008). The TACC proteins: TACC-ling microtubule dynamics and centrosome function. Trends Cell. Biol. *18*, 379-388.

Prensner, J.R., and Chinnaiyan, A.M. (2009). Oncogenic gene fusions in epithelial carcinomas. Curr Opin Genet. Dev. *19*, 82-91.

Reardon, D.A., Desjardins, A., Vredenburgh, J.J., Gururangan, S., Friedman, A.H., Herndon, J.E., 2nd, Marcello, J., Norfleet, J.A., McLendon, R.E., Sampson, J.H., *et al.* (2010). Phase 2

trial of erlotinib plus sirolimus in adults with recurrent glioblastoma. J. Neurooncol. *96*, 219-230.

Sheltzer, J.M., and Amon, A. (2011). The aneuploidy paradox: costs and benefits of an incorrect karyotype. Trends Genet. *27*, 446-453.

Solomon, D.A., Kim, T., Diaz-Martinez, L.A., Fair, J., Elkahloun, A.G., Harris, B.T., Toretsky, J.A., Rosenberg, S.A., Shukla, N., Ladanyi, M., *et al.* (2011). Mutational inactivation of STAG2 causes aneuploidy in human cancer. Science *333*, 1039-1043.

Stephens, P.J., McBride, D.J., Lin, M.L., Varela, I., Pleasance, E.D., Simpson, J.T., Stebbings, L.A., Leroy, C., Edkins, S., Mudie, L.J., *et al.* (2009). Complex landscapes of somatic rearrangement in human breast cancer genomes. Nature *462*, 1005-1010.

Still, I.H., Vince, P., and Cowell, J.K. (1999). The third member of the transforming acidic coiled coil-containing gene family, TACC3, maps in 4p16, close to translocation breakpoints in multiple myeloma, and is upregulated in various cancer cell lines. Genomics *58*, 165-170.

Thompson, S.L., Bakhoum, S.F., and Compton, D.A. (2010). Mechanisms of chromosomal instability. Curr. Biol. *20*, R285-295.

Tomlins, S.A., Laxman, B., Dhanasekaran, S.M., Helgeson, B.E., Cao, X., Morris, D.S., Menon, A., Jing, X., Cao, Q., Han, B., *et al.* (2007). Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. Nature *448*, 595-599.

Tomlins, S.A., Rhodes, D.R., Perner, S., Dhanasekaran, S.M., Mehra, R., Sun, X.W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R., *et al.* (2005). Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science *310*, 644-648.

Turner, N., and Grose, R. (2010). Fibroblast growth factor signalling: from development to cancer. Nat. Rev. Cancer *10*, 116-129.

van den Bent, M.J., Brandes, A.A., Rampling, R., Kouwenhoven, M.C., Kros, J.M., Carpentier, A.F., Clement, P.M., Frenay, M., Campone, M., Baurain, J.F., *et al.* (2009). Randomized phase II trial of erlotinib versus temozolomide or carmustine in recurrent glioblastoma: EORTC brain tumor group study 26034. J. Clin. Oncol. *27*, 1268-1274.

Wang, X.S., Prensner, J.R., Chen, G., Cao, Q., Han, B., Dhanasekaran, S.M., Ponnala, R., Cao, X., Varambally, S., Thomas, D.G., *et al.* (2009). An integrative approach to reveal driver gene fusions from paired-end sequencing data in cancer. Nat. Biotechnol. *27*, 1005-1011.

Weaver, B.A., and Cleveland, D.W. (2009). The role of aneuploidy in promoting and suppressing tumors. J. Cell. Biol. *185*, 935-937.

Wesche, J., Haglund, K., and Haugsten, E.M. (2011). Fibroblast growth factors and their receptors in cancer. Biochem. J. *437*, 199-213.

Yan, H., Parsons, D.W., Jin, G., McLendon, R., Rasheed, B.A., Yuan, W., Kos, I., Batinic-Haberle, I., Jones, S., Riggins, G.J., *et al.* (2009). IDH1 and IDH2 mutations in gliomas. New Engl. J. Med. *360*, 765-773.

Zhao, X., D, D.A., Lim, W.K., Brahmachary, M., Carro, M.S., Ludwig, T., Cardo, C.C., Guillemot, F., Aldape, K., Califano, A., *et al.* (2009). The N-Myc-DLL3 cascade is suppressed by the ubiquitin ligase Huwe1 to inhibit proliferation and promote neurogenesis in the developing brain. Dev. Cell *17*, 210-221.

Zhao, X., Heng, J.I., Guardavaccaro, D., Jiang, R., Pagano, M., Guillemot, F., Iavarone, A., and Lasorella, A. (2008). The HECT-domain ubiquitin ligase Huwe1 controls neural differentiation and proliferation by destabilizing the N-Myc oncoprotein. Nat Cell Biol *10*, 643-653.

(2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature *455*, 1061-1068.

Iglewicz, B., and Hoaglin, D.C. (1993). How to detect and handle outliers (Milwaukee, Wis.: ASQC).

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754-1760.

Morganella, S., Cerulo, L., Viglietto, G., and Ceccarelli, M. (2010). VEGA: variational segmentation for copy number detection. Bioinformatics *26*, 3020-3027.

Seal, R.L., Gordon, S.M., Lush, M.J., Wright, M.W., and Bruford, E.A. (2011). genenames.org: the HGNC resources in 2011. Nucleic Acids Res *39*, D514-519.

Tomlins, S.A., Rhodes, D.R., Perner, S., Dhanasekaran, S.M., Mehra, R., Sun, X.W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R., *et al.* (2005). Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science *310*, 644-648.

Vilella, A.J., Severin, J., Ureta-Vidal, A., Heng, L., Durbin, R., and Birney, E. (2009). EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. Genome Res. *19*, 327-335.

Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S.F., Hakonarson, H., and Bucan, M. (2007). PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. *17*, 1665-1674.

Wang, X.S., Prensner, J.R., Chen, G., Cao, Q., Han, B., Dhanasekaran, S.M., Ponnala, R., Cao, X., Varambally, S., Thomas, D.G., *et al.* (2009). An integrative approach to reveal driver gene fusions from paired-end sequencing data in cancer. Nat. Biotechnol. *27*, 1005-1011.

[00315] **Example 2-** ***Fusions in GBM***

[00316] **Table 8: Soft agar colony assay**

Cell line	Vector	FGFR3	TACC3	F1-T1 Fusion	F3-T3 Fusion	F3-T3-K508M Fusion
Rat1	0	0	0	225.3±10.0	198.7±8.0	0
Balb 3T3	0	0	0	n.d.	45.5±8.9	n.d.

n.d.: not done

[00317] Table 9: Subcutaneous tumor xenografts

Cell line	Vector	FGFR3	TACC3	F1-T1 Fusion	F3-T3 Fusion	F3-T3-K508M Fusion
Rat1	0/5	0/5	0/5	n.d.	5/5	n.d.
Ink4A/Arf-/- Astrocytes	0/9	0/5	0/5	8/8	12/12	0/8

n.d.: not done

[00318] Table 10: Analysis of chromosomal number in Rat1 cells

Cell line	Number of cells counted	Percent aneuploidy	Range	Mean number	Average variation from mean number	*p*- value
Rat1A Vector	100	27	35-43	41.2	1.2	
Rat1A FGFR3	100	33	35-44	42.1	1.3	n.s.
Rat1A TACC3	100	41	34-46	40.7	1.1	n.s.
Rat1A FGFR3-TACC3	100	69	35-73	43.8	3.1	<0.0001

[00319] Table 11: Analysis of chromosomal number in human astrocytes

Cell line	Number of cells counted	Percent aneuploidy	Range	Mean number	Average variation from mean number	*p*- value
Human Astrocytes Vector	100	8	42-46	45.85	0.28	$p = <0.001$
Human Astrocytes FGFR3-TACC3	100	42	28-48	42.24	3.33	

[00320] Example 3 - *Fusions in Other Cancers*

[00321] The inventors previously reported in Example 1 that 3.1% of human glioblastoma harbor FGFR3-TACC3 and FGFR1-TACC1 gene fusions. Tumors harboring FGFR3-TACC3 gene fusions are identified by the presence of highly specific focal micro-amplification events of the rearranged portions of the FGFR3 and TACC3 genes (*See* **FIG. 2E**). Therefore, these micro-amplification events can be used as distinctive marks for the presence of FGFR3-TACC3 gene fusions. It was asked whether other types of human tumors

also harbor FGFR3-TACC3 gene fusions from the analysis of Copy Number Variations (CNVs) of SNP arrays generated from the Atlas-TCGA project. This analysis was performed using segmented CNVs data visualized using the Integrated Genomic Viewers software. The analysis revealed that the following tumors, shown in the **FIGS. 31-35**, display focal micro-amplification events of FGFR3 and TACC3 that indicate the presence of FGFR3-TACC3 gene fusions (in **FIGS. 31-35**, red indicates amplification (A), blue indicates deletion (D); **FIG. 31**: Bladder Urothelial Carcinoma; **FIG. 32**: Breast Carcinoma; **FIG. 33**: Colorectal Carcinoma; **FIG. 34**: Lung Squamous Cell Carcinoma; **FIG. 35**: Head and Neck Squamous Cell Carcinoma).

[00322] Taken together, these data indicate that the same FGFR3-TACC3 gene fusions reported for the first time in Glioblastoma also occur in several other types of human tumors. Therefore, as for Glioblastoma and other epithelial cancers (such as the human tumors discussed herein), the identification of FGFR-TACC gene fusions also provides a new diagnostic and therapeutic target for treatment with drugs that inhibit FGFR-TACC gene fusions.

What is claimed is:

1. A purified fusion protein comprising a tyrosine kinase domain of an FGFR protein fused to a polypeptide that constitutively activates the tyrosine kinase domain of the FGFR protein.

2. A purified fusion protein comprising a transforming acidic coiled-coil (TACC) domain fused to a polypeptide with a tyrosine kinase domain, wherein the TACC domain constitutively activates the tyrosine kinase domain.

3. A purified fusion protein comprising the tyrosine kinase domain of an FGFR protein fused 5' to the TACC domain of a transforming acidic coiled-coil-containing (TACC) protein.

4. The purified fusion protein of claim 1 or 3, wherein the FGFR protein is FGFR1, FGFR2, FGFR3, or FGR4.

5. The purified fusion protein of claim 3, wherein the TACC protein is TACC1, TACC2, or TACC3.

6. A purified fusion protein encoded by an FGFR1-TACC1 nucleic acid, wherein FGFR1-TACC1 comprises a combination of exons 1-17 of FGFR1 located on human chromosome 8p11 spliced 5' to a combination of exons 7-13 of TACC1 located on human chromosome 8p11, wherein a genomic breakpoint occurs in any one of exons 1-17 of FGFR1 and any one of exons 7-13 of TACC1.

7. A purified fusion protein encoded by an FGFR2-TACC2 nucleic acid, wherein FGFR2-TACC2 comprises a combination of any exons 1-18 of FGFR2 located on human chromosome 10q26 spliced 5' to a combination of any exons 1-23 of TACC2 located on human chromosome 10q26.

8. A purified fusion protein encoded by an FGFR3-TACC3 nucleic acid, wherein FGFR3-TACC3 comprises a combination of exons 1-16 of FGFR3 located on human chromosome 4p16 spliced 5' to a combination of exons 8-16 of TACC3 located on human chromosome 4p16, wherein a genomic breakpoint occurs in any one of exons 1-16 of FGFR3 and any one of exons 8-16 of TACC3.

9. A purified fusion protein encoded by an FGFR3-TACC3 nucleic acid, wherein FGFR3-TACC3 comprises a combination of exons 1-16 of FGFR3 located on human chromosome 4p16 spliced 5' to a combination of exons 8-16 of TACC3 located on human chromosome 4p16, wherein a genomic breakpoint occurs in any one of *introns* 1-16 of FGFR3 and any one of exons 8-16 of TACC3.

10. A purified fusion protein encoded by an FGFR3-TACC3 nucleic acid, wherein FGFR3-TACC3 comprises a combination of exons 1-16 of FGFR3 located on human chromosome 4p16 spliced 5' to a combination of exons 8-16 of TACC3 located on human chromosome 4p16, wherein a genomic breakpoint occurs in any one of exons 1-16 of FGFR3 and any one of *introns* 7-16 of TACC3.

11. A purified fusion protein encoded by an FGFR3-TACC3 nucleic acid, wherein FGFR3-TACC3 comprises a combination of exons 1-16 of FGFR3 located on human chromosome 4p16 spliced 5' to a combination of exons 8-16 of TACC3 located on human chromosome 4p16, wherein a genomic breakpoint occurs in any one of *introns* 1-16 of FGFR3 and any one of *introns* 7-16 of TACC3.

12. A synthetic nucleic acid encoding the fusion protein of claim 1, 2, 3, 6, 7, 8, 9, 10, or 11.

13. A purified FGFR3-TACC3 fusion protein comprising SEQ ID NO: 79, 158, 159, 160,

or 161.

14. A purified FGFR3-TACC3 fusion protein having a genomic breakpoint comprising at least 3 consecutive amino acids from amino acids 730-758 of SEQ ID NO: 90 and comprising at least 3 consecutive amino acids from amino acids 549-838 of SEQ ID NO: 92.

15. A purified FGFR3-TACC3 fusion protein having a genomic breakpoint comprising SEQ ID NO: 78.

16. A purified FGFR3-TACC3 fusion protein having a genomic breakpoint comprising any one of SEQ ID NOS: 85, 86, 87, or 89.

17. A purified FGFR1-TACC1 fusion protein comprising SEQ ID NO: 150.

18. A purified FGFR1-TACC1 fusion protein having a genomic breakpoint comprising at least 3 consecutive amino acids from amino acids 746-762 of SEQ ID NO: 146 and comprising at least 3 consecutive amino acids from amino acids 572-590 of SEQ ID NO: 148.

19. A purified FGFR1-TACC1 fusion protein having a genomic breakpoint comprising SEQ ID NO: 88.

20. A synthetic nucleic acid encoding an FGFR3-TACC3 fusion protein comprising SEQ ID NO: 94.

21. A synthetic nucleic acid encoding an FGFR3-TACC3 fusion protein having a genomic breakpoint comprising at least 9 consecutive in-frame nucleotides from nucleotides 2443-2530 of SEQ ID NO: 91 and comprising at least 9 consecutive in-frame nucleotides from nucleotides 1800-2847 of SEQ ID NO: 93.

22. A synthetic nucleic acid encoding an FGFR3-TACC3 fusion protein having a genomic breakpoint comprising any one of SEQ ID NOS: 1-77.

23. A synthetic nucleic acid encoding an FGFR1-TACC1 fusion protein comprising SEQ ID NO: 151.

24. A synthetic nucleic acid encoding an FGFR1-TACC1 fusion protein having a genomic breakpoint comprising at least 9 consecutive in-frame nucleotides from nucleotides 3178-3228 of SEQ ID NO: 147 and comprising at least 9 consecutive in-frame nucleotides from nucleotides 2092-2794 of SEQ ID NO: 149.

25. A synthetic nucleic acid encoding an FGFR1-TACC1 fusion protein having a genomic breakpoint comprising SEQ ID NO: 83.

26. An antibody or antigen-binding fragment thereof, that specifically binds to a purified fusion protein comprising a tyrosine kinase domain of an FGFR protein fused to a polypeptide that constitutively activates the tyrosine kinase domain of the FGFR protein.

27. The antibody or antigen-binding fragment of claim 26, wherein the FGFR protein is FGFR1, FGFR2, FGFR3, or FGFR4.

28. The antibody or antigen-binding fragment of claim 26, wherein the fusion protein is an FGFR-TACC fusion protein.

29. The antibody or antigen-binding fragment of claim 28, wherein the FGFR-TACC fusion protein is FGFR1-TACC1, FGFR2-TACC2, or FGFR3-TACC3.

30. The antibody or antigen-binding fragment of claim 29, wherein the FGFR1-TACC1 fusion protein comprises the amino acid sequence of SEQ ID NO: 150.

31. The antibody or antigen-binding fragment of claim 29, wherein the FGFR3-TACC3 fusion protein comprises the amino acid sequence of SEQ ID NO: 79, 158, 159, 160, or 161.

32. A composition for decreasing in a subject the expression level or activity of a fusion protein comprising the tyrosine kinase domain of an FGFR protein fused to a polypeptide that constitutively activates the tyrosine kinase domain of the FGFR protein, the composition in an admixture of a pharmaceutically acceptable carrier comprising an inhibitor of the fusion protein.

33. The composition of claim 32, wherein the fusion protein is an FGFR-TACC fusion protein.

34. The composition of claim 32, wherein the inhibitor comprises an antibody that specifically binds to a FGFR-TACC fusion protein or a fragment thereof; a small molecule that specifically binds to a FGFR protein; a small molecule that specifically binds to a TACC protein; an antisense RNA or antisense DNA that decreases expression of a FGFR-TACC fusion polypeptide; a siRNA that specifically targets a FGFR-TACC fusion gene; or a combination thereof.

35. The composition of claim 32, wherein the FGFR protein is FGFR1, FGFR2, FGFR3, or FGFR4.

36. The composition of claim 33, wherein the FGFR-TACC fusion protein is FGFR1-TACC1, FGFR2-TACC2, or FGFR3-TACC3.

37. The composition of claim 34, wherein the small molecule that specifically binds to a FGFR protein comprises AZD4547, NVP-BGJ398, PD173074, NF449, TK1258, BIBF-1120, BMS-582664, AZD-2171, TSU68, AB1010, AP24534, E-7080,

LY2874455, or a combination thereof.

38. A method for decreasing in a subject in need thereof the expression level or activity of a fusion protein comprising the tyrosine kinase domain of an FGFR protein fused to a polypeptide that constitutively activates the tyrosine kinase domain of the FGFR protein, the method comprising:

(a) administering to the subject a therapeutic amount of a composition of claim 32; and

(b) determining whether the fusion protein expression level or activity is decreased compared to fusion protein expression level or activity prior to administration of the composition,

thereby decreasing the expression level or activity of the fusion protein.

39. A method for treating a gene-fusion associated cancer in a subject in need thereof, the method comprising administering to the subject an effective amount of a FGFR fusion molecule inhibitor.

40. A method of decreasing growth of a solid tumor in a subject in need thereof, the method comprising administering to the subject an effective amount of a FGFR fusion molecule inhibitor, wherein the inhibitor decreases the size of the solid tumor.

41. The method of claim 39, wherein the gene-fusion associated cancer comprises glioblastoma multiforme, breast cancer, lung cancer, prostate cancer, or colorectal carcinoma.

42. The method of claim 40, wherein the solid tumor comprises glioblastoma multiforme, breast cancer, lung cancer, prostate cancer, or colorectal carcinoma.

43. The method of claim 39 or 40, wherein the FGFR fusion protein comprises an FGFR protein fused to a polypeptide that constitutively activates the tyrosine kinase domain of the FGFR protein.

44. The method of claim 38, 39 or 40, wherein the fusion protein is an FGFR-TACC fusion protein.

45. The method of claim 39 or 40, wherein the inhibitor comprises an antibody that specifically binds to a FGFR-TACC fusion protein or a fragment thereof; a small molecule that specifically binds to a FGFR protein; a small molecule that specifically binds to a TACC protein; an antisense RNA or antisense DNA that decreases expression of a FGFR-TACC fusion polypeptide; a siRNA that specifically targets a FGFR-TACC fusion gene; or a combination thereof.

46. The method of claim 38 or 43, wherein the FGFR protein is FGFR1, FGFR2, FGFR3, or FGFR4.

47. The method of claim 45, wherein the FGFR-TACC fusion protein is FGFR1-TACC1, FGFR2-TACC2, or FGFR3-TACC3.

48. The method of claim 45, wherein the small molecule that specifically binds to a FGFR protein comprises AZD4547, NVP-BGJ398, PD173074, NF449, TK1258, BIBF-1120, BMS-582664, AZD-2171, TSU68, AB1010, AP24534, E-7080, LY2874455, or a combination thereof.

49. A diagnostic kit for determining whether a sample from a subject exhibits a presence of a FGFR fusion, the kit comprising at least one oligonucleotide that specifically hybridizes to a FGFR fusion, or a portion thereof.

50. The kit of claim 49, wherein the oligonucleotides comprise a set of nucleic acid primers or *in situ* hybridization probes.

51. The kit of claim 49, wherein the oligonucleotide comprises SEQ ID NO: 162, 163, 164,165, 166, 167, 168, 169, or a combination thereof.

52. The kit of claim 50, wherein the primers prime a polymerase reaction only when a FGFR fusion is present.

53. The kit of claim 49, wherein the determining comprises gene sequencing, selective hybridization, selective amplification, gene expression analysis, or a combination thereof.

54. A diagnostic kit for determining whether a sample from a subject exhibits a presence of a FGFR fusion protein, the kit comprising an antibody that specifically binds to a FGFR fusion protein comprising SEQ ID NO: 79, 85, 86, 87, 88, 89, 150, 158, 159, 160, or 161, wherein the antibody will recognize the protein only when a FGFR fusion protein is present.

55. The kit of claim 54, wherein the antibody directed to and FGFR fusion comprising SEQ ID NO: 79, 85, 86, 87, 88, 89, 150, 158, 159, 160, or 161.

56. The kit of claim 49 or 54, wherein the FGFR-fusion is an FGFR-TACC fusion.

57. The kit of claim 49 or 54, wherein the FGFR is FGFR1, FGFR2, FGFR3, or FGFR4.

58. The kit of claim 55, wherein the FGFR-TACC fusion is FGFR1-TACC1, FGFR2-TACC2, or FGFR3-TACC3.

59. A method for detecting the presence of a FGFR fusion in a human subject, the method

comprising:

(a) obtaining a biological sample from the human subject; and

(b) detecting whether or not there is a FGFR fusion present in the subject.

60. The method of claim 59, wherein the detecting comprises measuring FGFR fusion protein levels by ELISA using an antibody directed to SEQ ID NO: 79, 85, 86, 87, 88, 89, 150, 158, 159, 160, or 161; western blot using an antibody directed to SEQ ID NO: 79, 85, 86, 87, 88, 89, 150, 158, 159, 160, or 161; mass spectroscopy, isoelectric focusing, or a combination thereof.

61. A method for detecting the presence of a FGFR fusion in a human subject, the method comprising:

(a) obtaining a biological sample from a human subject; and

(b) detecting whether or not there is a nucleic acid sequence encoding a FGFR fusion protein in the subject.

62. The method of claim 61, wherein the nucleic acid sequence comprises any one of SEQ ID NOS: 1-77, 80-84, or 95-145.

63. The method of claim 61, wherein the detecting comprises using hybridization, amplification, or sequencing techniques to detect a FGFR fusion.

64. The method of claim 63, wherein the amplification uses primers comprising SEQ ID NO: 162, 163, 164,165, 166, 167, 168, or 169.

65. The method of claim 59 or 61, wherein the FGFR-fusion is an FGFR-TACC fusion.

66. The method of claim 59 or 61, wherein the FGFR is FGFR1, FGFR2, FGFR3, or

FGFR4.

67. The method of claim 65, wherein the FGFR-TACC fusion is FGFR1-TACC1, FGFR2-TACC2, or FGFR3-TACC3.

68. A method of initiating oncogenic transformation *in vitro*, the method comprising:

(a) transducing cells cultured *in vitro* with FGFR-TACC fusion DNA; and

(b) determining whether the cells acquire the ability to grow in anchorage-independent conditions, form multi-layered foci, or a combination thereof.

69. A method of initiating oncogenic transformation *in vivo*, the method comprising:

(a) transducing cells cultured *in vitro* with FGFR-TACC fusion DNA;

(b) injecting a mouse with the transduced cells; and

(c) determining whether a tumor grows in the mouse.

70. The method of claim 69, wherein the injecting is a subcutaneous or intracranial injection.

71. A method of identifying a compound that decreases the oncogenic activity of a FGFR-TACC fusion, the method comprising:

(a) transducing a cell cultured *in vitro* with FGFR-TACC DNA;

(b) contacting a cell with a ligand source for an effective period of time; and

(c) determining whether the cells acquire the ability to grow in anchorage-independent conditions, form multi-layered foci, or a combination

thereof, compared to cells cultured in the absence of the test compound.

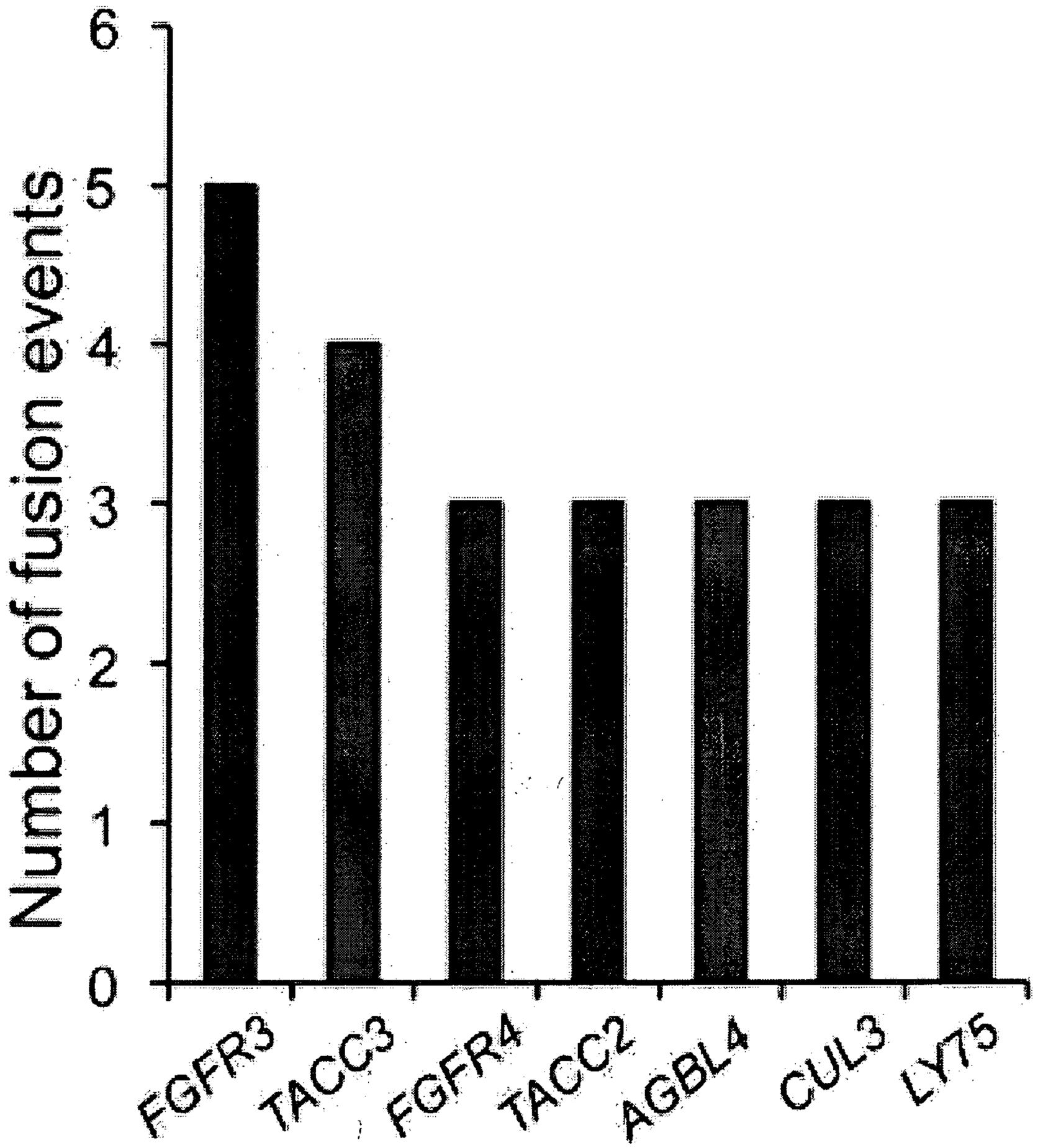

Number of fusion events
0
1
2
3
4
5
6
FGFR3
TACC3
FGFR4
TACC2
AGBL4
CUL3
LY75

FIG. 1A

NM_000142
FGFR3

5'

Reference: C W H A A P S Q R P T F K Q L V E D L D R V L T V T S

GTGCTGGCATGCCGCGCCCTCCCAGAGGCCCACCTTCAAGCAGCTGGTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC

SpiltReads:

C
TC
GTC
CGTC
GACGTC
TGACGTC
GTGACGTC
CGTGACGTC
CCGTGACGTC
CTTACCGTGACGTC
GTCCTTACCGTGACGTC
TGTCCTTACCGTGACGTC
GTGTCCTTACCGTGACGTC
CGTGTCCTTACCGTGACGTC
ACCGTGTCCTTACCGTGACGTC
GACCGTGTCCTTACCGTGACGTC
GGACCGTGTCCTTACCGTGACGTC
CTGGACCGTGTCCTTACCGTGACGTC
CCTGGACCGTGTCCTTACCGTGACGTC
GACCTGGACCGTGTCCTTACCGTGACGTC
GGACCTGGACCGTGTCCTTACCGTGACGTC
AGGACCTGGACCGTGTCCTTACCGTGACGTC
GAGGACCTGGACCGTGTCCTTACCGTGACGTC
GGAGGACCTGGACCGTGTCCTTACCGTGACGTC
TGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
GTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
GGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
TGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC

Cont. on FIG.1B-2

FIG.1B-1

Cont. on FIG.1B-3

NM_006342
TACC3

2530 1751 3'

T D F K E S A L R K Q S L Y L K F D P L L R D S P G R P V P V A

CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCCCTCCTGAGGGACAGTCCTGGTAGACCAGTGCCCGTGGCC

C TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCCCTCCTGAGGGACAGTCCTGGTAGACCAGTGCCCGTGGCC
AC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCCCTCCTGAGGGACAGTCCTGGTAGACCAGTGCCCGTGGG
GAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCCCTCCTGAGGGACAGTCCTGGTAGACCAGTGCCCGTGG
CGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCCCTCCTGAGGGACAGTCCTGGTAGACCAGTGCCCGTT
CCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCCCTCCTGAGGGACAGTCCTGGTAGACCAGTGCCCGG
ACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCCCTCCTGAGGGACAGTCCTGGTAGACCAGTGCCCC
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCCCTCCTGAGGGACAGTCCTGGTAGACCAGTGCCC
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCCCTCCTGAGGGACAGTCCTGGTAGACCAGTGCC
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCCCTCCTGAGGGACAGTCCTGGTAGACCAGTGG
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCCCTCCTGAGGGACAGTCCTGGTAGACCAGTT
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCCCTCCTGAGGGACAGTCCTGGTAGACCAGG
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCCCTCCTGAGGGACAGTCCTGGTAGACG
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCCCTCCTGAGGGACAGTCCTGGTAGACC
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCCCTCCTGAGGGACAGTCCTGGTAGAA
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCCCTCCTGAGGGACAGTCCTGGTAGG
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCCCTCCTGAGGGACAGTCCTGGTAA
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCCCTCCTGAGGGACAGTCCTT
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCCCTCCTGAGGGACAGTT
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCCCTCCTGAGGGACAGG
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCCCTCCTGAGGGACAA
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCCCTCCTGAGGGACC
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCCCTCCTGAGGGG
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCCCTCCTGAGGG
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCCCTCCTGAGG
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCCCTCCTGG
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCCCTCCTT
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCCCTCC
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCCCTT
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCCCC
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCCC
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACCC
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGACC
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGAC
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCGA
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTTCG

Cont. from FIG.1B-1

FIG.1B-2

Cont. on FIG.1B-4

Cont. FROM FIG.1B-1

CTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
GCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
AGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
CAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
GCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
AGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
AAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
CAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
TCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
TTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
CTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
CCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
ACCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
CACCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
CCACCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
CCCACCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
GGCCCACCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
AGGCCCACCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
GAGGCCCACCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
AGAGGCCCACCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
CAGAGGCCCACCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
CCAGAGGCCCACCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
CCCAGAGGCCCACCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
TCCCAGAGGCCCACCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
CTCCCAGAGGCCCACCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
CCTCCCAGAGGCCCACCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
CCCTCCCAGAGGCCCACCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
GCGCCCTCCCAGAGGCCCACCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
CGCGCCCTCCCAGAGGCCCACCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
CCGCGCCCTCCCAGAGGCCCACCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
GCCGCGCCCTCCCAGAGGCCCACCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
TGCCGCGCCCTCCCAGAGGCCCACCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
ATGCCGCGCCCTCCCAGAGGCCCACCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
CATGCCGCGCCCTCCCAGAGGCCCACCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
GCATGCCGCGCCCTCCCAGAGGCCCACCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
GGCATGCCGCGCCCTCCCAGAGGCCCACCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
TGGCATGCCGCGCCCTCCCAGAGGCCCACCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
CTGGCATGCCGCGCCCTCCCAGAGGCCCACCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
GCTGGCATGCCGCGCCCTCCCAGAGGCCCACCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
TGCTGGCATGCCGCGCCCTCCCAGAGGCCCACCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC
GTGCTGGCATGCCGCGCCCTCCCAGAGGCCCACCTTCAAGCAGCTGGTGGAGGACCTGGACCGTGTCCTTACCGTGACGTC

Cont. on FIG.1B-4

FIG.1B-3

Cont. FROM FIG.1B-2

CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCTAATCA
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGTT
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAGT
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAAG
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCAA
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCTCA
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCCC
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACCT
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATACC
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATAC
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTATA
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTAT
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTTA
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCTT
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCCT
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGTCC
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAGT
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCAG
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGCA
Cont. from FIG.1B-3
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAGC
CACCGAC TTTAAGGAGTCGGCCTTGAGGAAG
CACCGAC TTTAAGGAGTCGGCCTTGAGGAA
CACCGAC TTTAAGGAGTCGGCCTTGAGGA
CACCGAC TTTAAGGAGTCGGCCTTGAGG
CACCGAC TTTAAGGAGTCGGCCTTGAG
CACCGAC TTTAAGGAGTCGGCCTTGA
CACCGAC TTTAAGGAGTCGGCCTTG
CACCGAC TTTAAGGAGTCGGCC
CACCGAC TTTAAGGAGTCGGC
CACCGAC TTTAAGGAGTCGG
CACCGAC TTTAAGGAGTCG
CACCGAC TTTAAGGAGTC
CACCGAC TTTAAGGAGT
CACCGAC TTTAAGGAG
CACCGAC TTTAAGGA
CACCGAC TTTAAGG
CACCGAC TTTAAG
CACCGAC TTTAA
CACCGAC TTTA
CACCGAC TTT
CACCGAC TT

FIG.1B-4

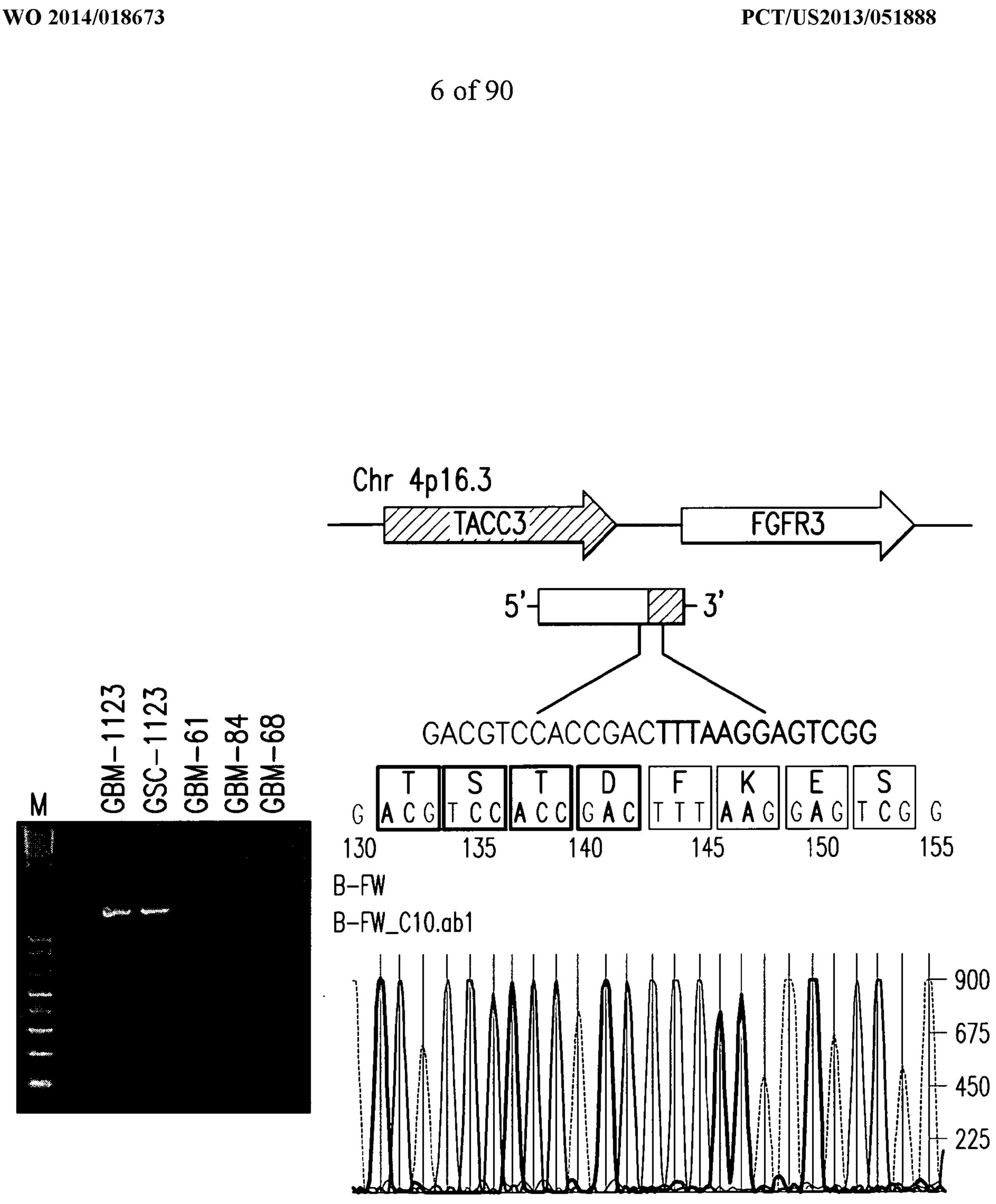

Chr 4p16.3
TACC3
FGFR3
5'
3'
GACGTCCACCGACTTTAAGGAGTCGG
T S T D F K E S
G A C G T C C A C C G A C T T T A A G G A G T C G G
130 135 140 145 150 155
B-FW
B-FW_C10.ab1
900
675
450
225
M
GBM-1123
GSC-1123
GBM-61
GBM-84
GBM-68

FIG.1C

MGAPACALALCVAVAIVAGASSESLGTEQRVVGRAAEVPGPEPGQQEQLVFGSGDAVELSCPPPG
GGPMGPTVWVKDGTGLVPSERVLVGPQRLQVLNASHEDSGAYSCRQRLTQRVLCHFSVRVTDAPS
SGDDEDGEDEAEDTGVDTGAPYWTRPERMDKKLLAVPAANTVRFRCPAAGNPTPSISWLKNGREF
RGEHRIGGIKLRHQQWSLVMESVVPSDRGNYTCVVENKFGSIRQTYTLDVLERSPHRPILQAGLP
ANQTAVLGSDVEFHCKVYSDAQPHIQWLKHVEVNGSKVGPDGTPYVTVLKTAGANTTDKELEVLS
LHNVTFEDAGEYTCLAGNSIGFSHHSAWLVVLPAEEELVEADEAGSVYAGILSYGVGFFLFILVV
AAVTLCRLRSPPKKGLGSPTVHKISRFPLKRQVSLESNASMSSNTPLVRIARLSSGEGPTLANVS
ELELPADPKWELSRARLTLGKPLGEGCFGQVVMAEAIGIDKDRAAKPVTVAVKMLKDDATDKDLS
DLVSEMEMMKMIGKHKNIINLLGACTQGGPLYVLVEYAAKGNLREFLRARRPPGLDYSFDTCKPP
EEQLTFKDLVSCAYQVARGMEYLASQKCIHRDLAARNVLVTEDNVMKIADFGLARDVHNLDYYKK
TTNGRLPVKWMAPEALFDRVYTHQSDVWSFGVLLWEIFTLGGSPYPGIPVEELFKLLKEGHRMDK
PANCTHDLYMIMRECWHAAPSQRPTFKQLVEDLDRVLTVTSTDFKESALRKQSLYLKFDPLLRDS
PGRPVPVATETSSMHGANETPSGRPREAKLVEFDFLGALDIPVPGPPPGVPAPGGPPLSTGPIVD
LLQYSQKDLDAVVKATQEENRELRSRCEELHGKNLELGKIMDRFEEVVYQAMEEVQKQKELSKAE
IQKVLKEKDQLTTDLNSMEKSFSDLFKRFEKQKEVIEGYRKNEESLKKCVEDYLARITQEGQRYQ
ALKAHAEEKLQLANEEIAQVRSKAQAEALALQASLRKEQMRIQSLEKTVEQKTKENEELTRICDD
LISKMEKI

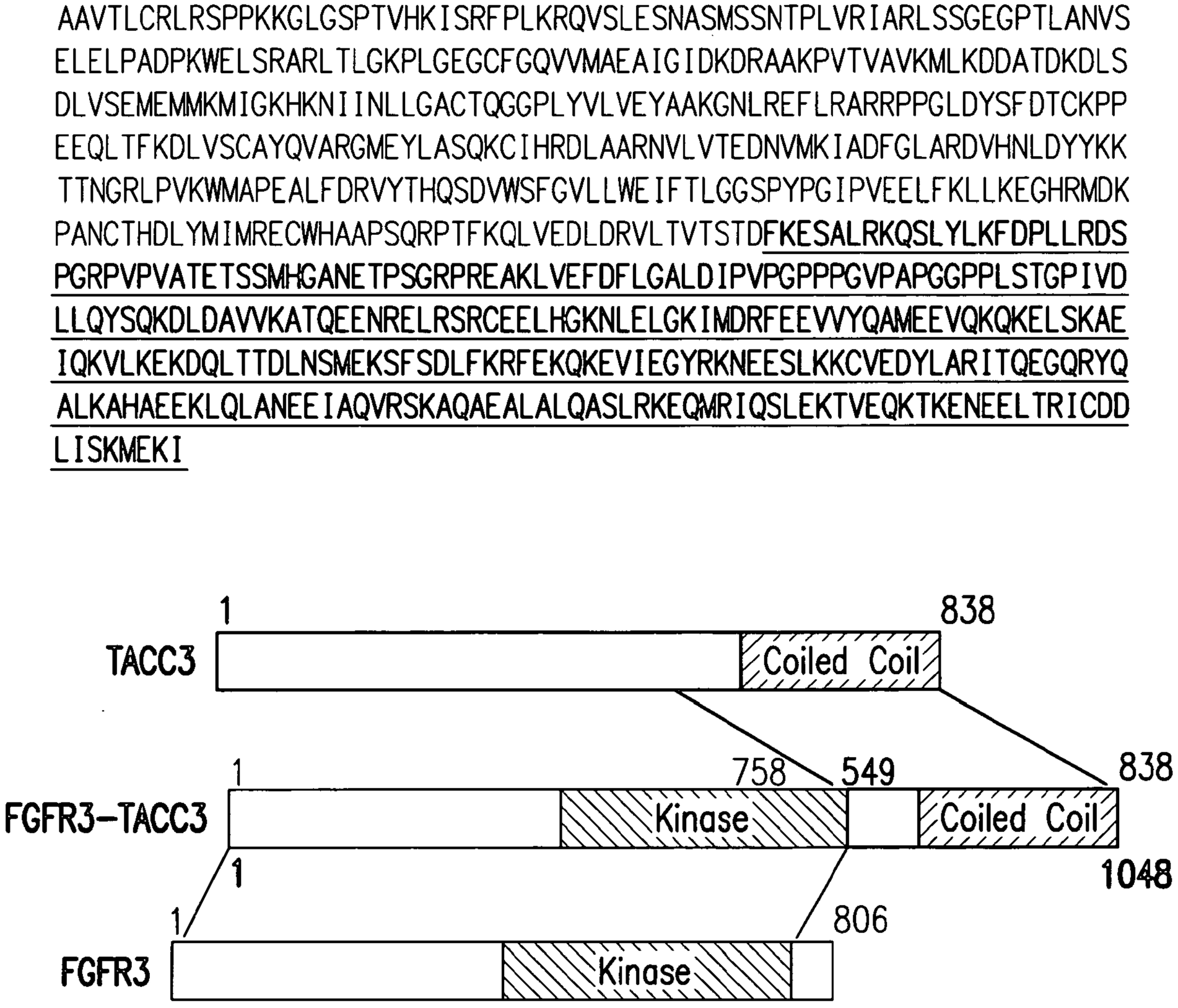


FIG.1D

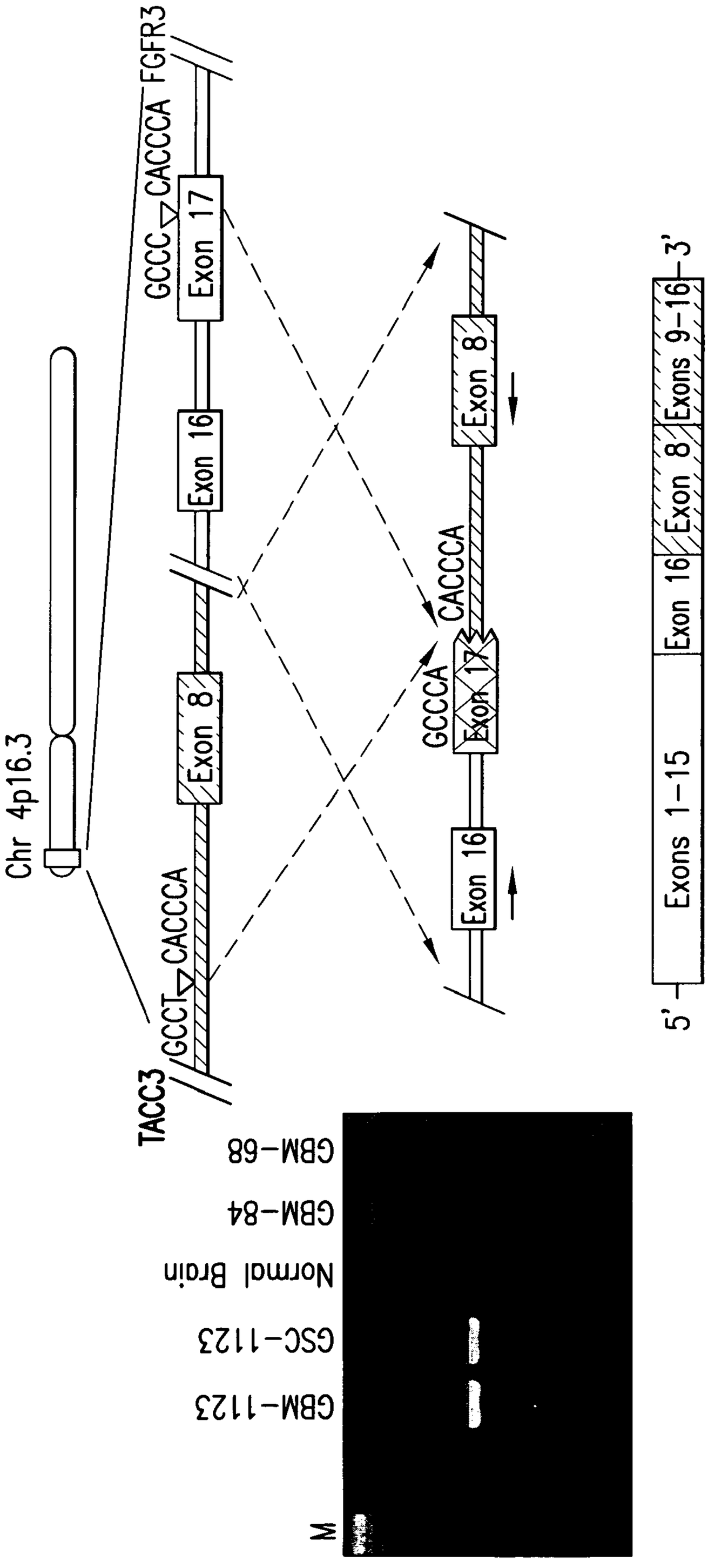

Chr 4p16.3
FGFR3
GCCC
CACCCA
Exon 17
Exon 16
Exon 8
GCCT
CACCCA
Exon 8
CACCCA
GCCCA
Exon 17
Exon 16
Exons 1–15
Exon 16
Exon 8
Exons 9–16
3'
5'
TACC3
GBM-68
GBM-84
Normal Brain
GSC-1123
GBM-1123
M

FIG.1E

TCGA-27-1835

5'	FGFR3: Intron 16	1778595	1709397	TACC3: Intron 10	3'
	CATCCTGCCCCCCAGAGTGCTGAGGTGTGGGGCGGGCCTTCTGGGGCACAGCCTGGGCACAGAGGTGGCTGTGCGA		AGGTCGCTGAGGGTCCAGGCTTCCACCCAGTGTCCCCGCAGTCAGCTGCCCACCAGCAGCCTCCCCGGGACTCTCC		
R	GGCTGTGCGA		AGGTCGCTGAGGGTCCAGGCTTCCACCCAGTGTCCCCGCAGTCAGCTGCCCACCAGCAGCCTCCCCC		
F	ACAGCCTGGGCACAGAGGTGGCTGTGCGA		AGGTCGCTGAGGGTCCAGGCTTCCACCCAGTGTCCCCGCAGTCAGCTT		
R	GGCACAGCCTGGGCACAGAGGTGGCTGTGCGA		AGGTCGCTGAGGGTCCAGGCTTCCACCCAGTGTCCCCGCAGTCAA		
R	TTCTGGGGCACAGCCTGGGCACAGAGGTGGCTGTGCGA		AGGTCGCTGAGGGTCCAGGCTTCCACCCAGTGTCCCCGG		
R	CTTCTGGGGCACAGCCTGGGCACAGAGGTGGCTGTGCGA		AGGTCGCTGAGGGTCCAGGCTTCCACCCAGTGTCCCCC		
R	GCCTTCTGGGGCACAGCCTGGGCACAGAGGTGGCTGTGCGA		AGGTCGCTGAGGGTCCAGGCTTCCACCCAGTGTCCC		
R	GGTGTGGGGCGGGCCTTCTGGGGCACAGCCTGGGCACAGAGGTGGCTGTGCGA		AGGTCGCTGAGGGTCCAGGCTTCC		
R	GGTGTGGGGCGGGCCTTCTGGGGCACAGCCTGGGCACAGAGGTGGCTGTGCGA		AGGTCGCTGAGGGTCCAGGCTTCC		
R	TGCTGAGGTGTGGGGCGGGCCTTCTGGGGCACAGCCTGGGCACAGAGGTGGCTGTGCGA		AGGTCGCTGAGGGTCCAA		
R	AGTGCTGAGGTGTGGGGCGGGCCTTCTGGGGCACAGCCTGGGCACAGAGGTGGCTGTGCGA		AGGTCGCTGAGGGTCC		
R	CCCAGAGTGCTGAGGTGTGGGGCGGGCCTTCTGGGGCACAGCCTGGGCACAGAGGTGGCTGTGCGA		AGGTCGCTGAA		
R	CCTGCCCCCCAGAGTGCTGAGGTGTGGGGCGGGCCTTCTGGGGCACAGCCTGGGCACAGAGGTGGCTGTGCGA		AGG		

TCGA-19-5958

5'	FGFR3: Intron 16	1778539	1707202	TACC3: Intron 7	3'
	GTGCTGGCTCTGGCCTGGTGCCACCCGCCTATGCCCCTCCCCCTGCCGTCCCCGGCCATCCTGCCCCCCCAGAGTGC		CGGGGGCTAAGGGCCAGGGAGGTCACCTGCACACTCCCACCCCCGGTCACCCGCACACTCCCACCCCCGGTCACCCA		
R	AGAGTGC		AGGGGGCTAAGGGCCAGGGAGGTCACCTGCACACTCCCACCCCCGGTCACCCGCACACTCCCACCCCCG		
R	GGCCATCCTGCCCCCCAGAGTGC		AGGGGGCTAAGGGCCAGGGAGGTCACCTGCACACTCCCACCCCCGGTCACCCG		
R	TCCCCCTGCCGTCCCCGGCCATCCTGCCCCCCAGAGTGC		AGGGGGCTAAGGGCCAGGGAGGTCACCTGCACACTCC		

FIG.2A

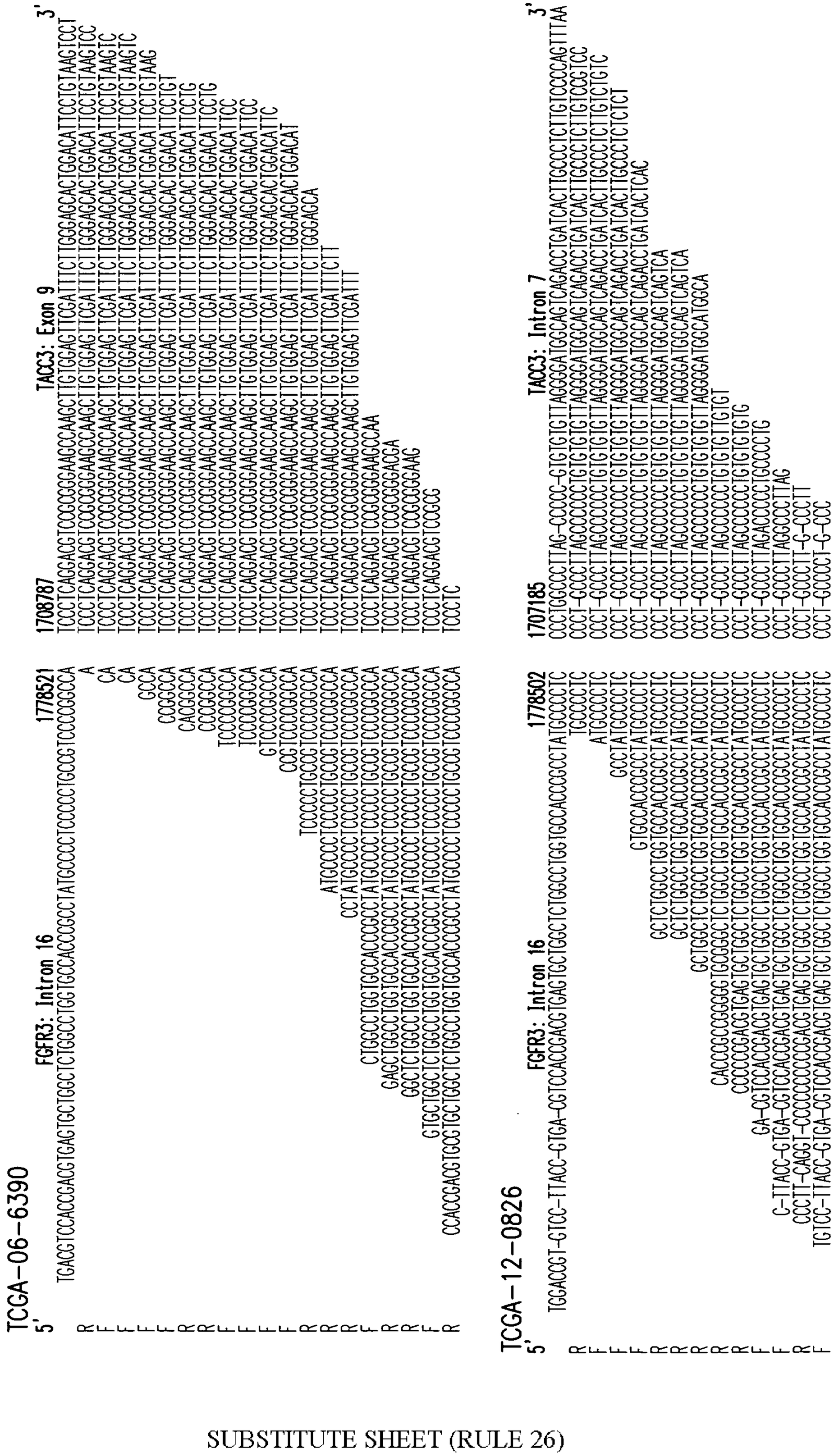

TACC3: Exon 9
TACC3: Intron 7
TACC3: Intron 16
FGFR3: Intron 16
TCGA-06-6390
TCGA-12-0826

FIG. 2A-CONT.

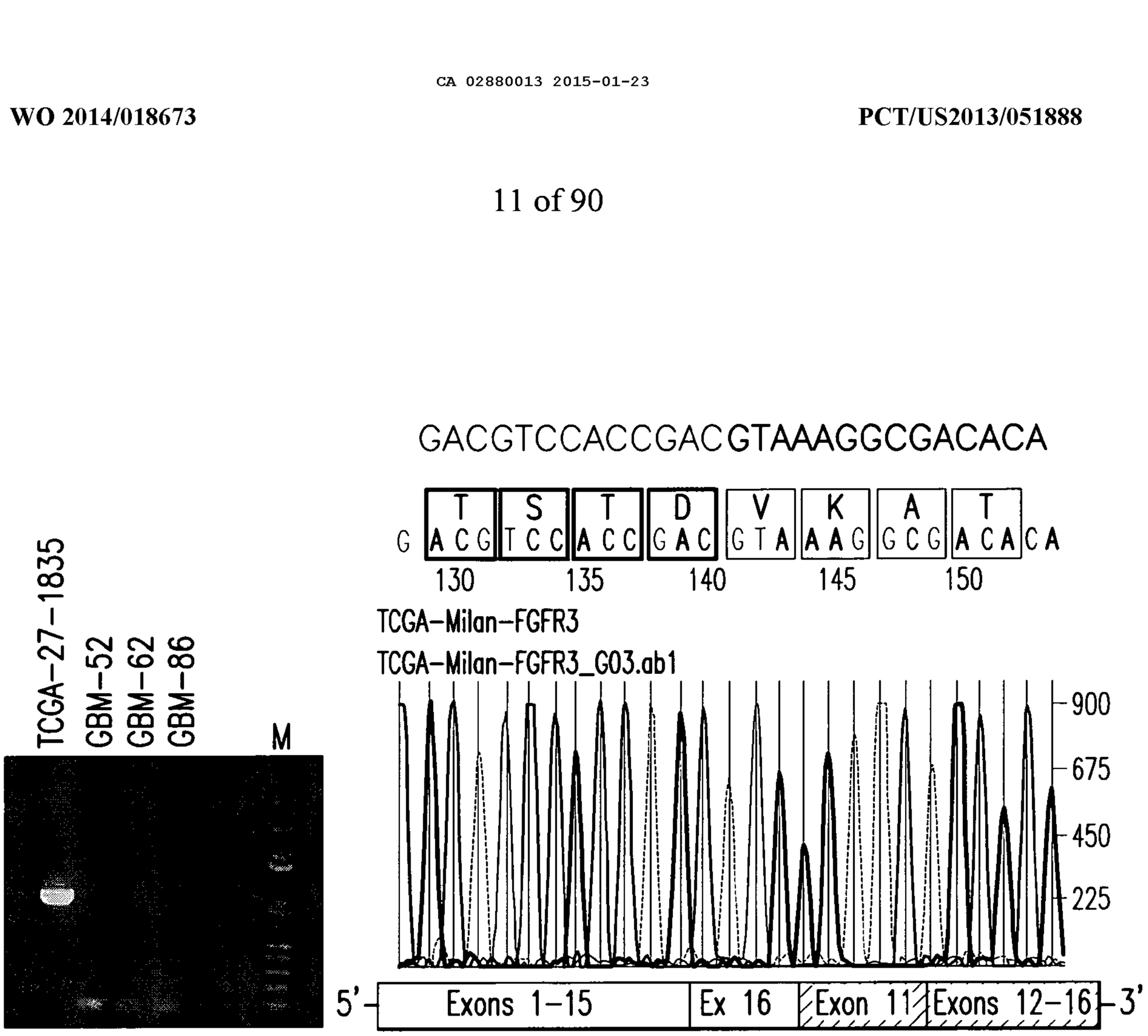

GACGTCCACCGACGTAAAGGCGACACA
T S T D V K A T
G A C G T C C A C C G A C G T A A A G G C G A C A C A
130 135 140 145 150
TCGA-Milan-FGFR3
TCGA-Milan-FGFR3_G03.ab1
900
675
450
225
TCGA-27-1835
GBM-52
GBM-62
GBM-86
M
5'
Exons 1-15
Ex 16
Exon 11
Exons 12-16
3'

FIG.2B

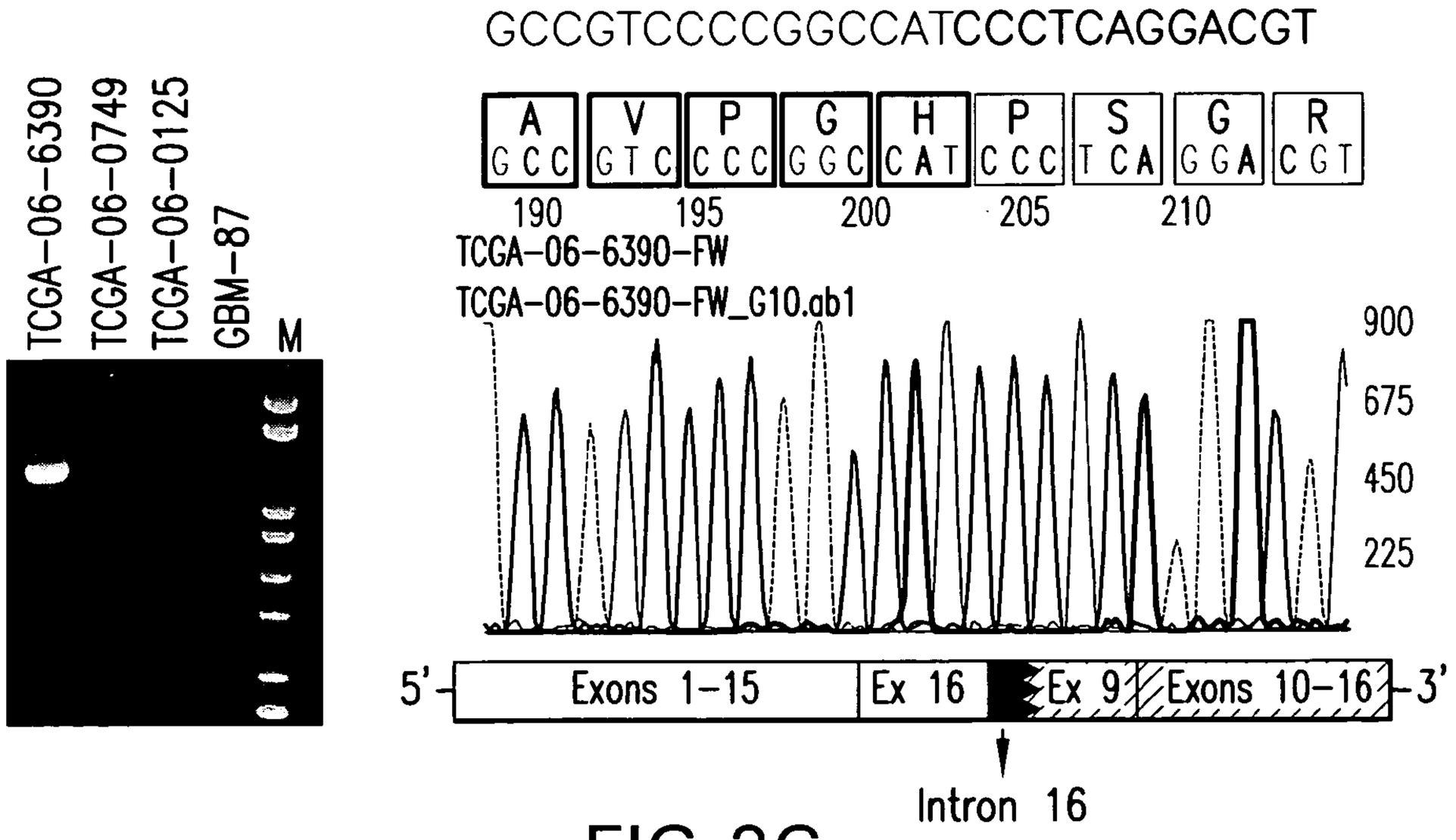

GCCGTCCCCGGCCATCCCTCAGGACGT
A V P G H P S G R
G C C G T C C C C C G G C C A T C C C T C A G G A C G T
190 195 200 205 210
TCGA-06-6390-FW
TCGA-06-6390-FW_G10.ab1
900
675
450
225
TCGA-06-6390
TCGA-06-0749
TCGA-06-0125
GBM-87
M
5'
Exons 1-15
Ex 16
Ex 9
Exons 10-16
3'
Intron 16

FIG.2C

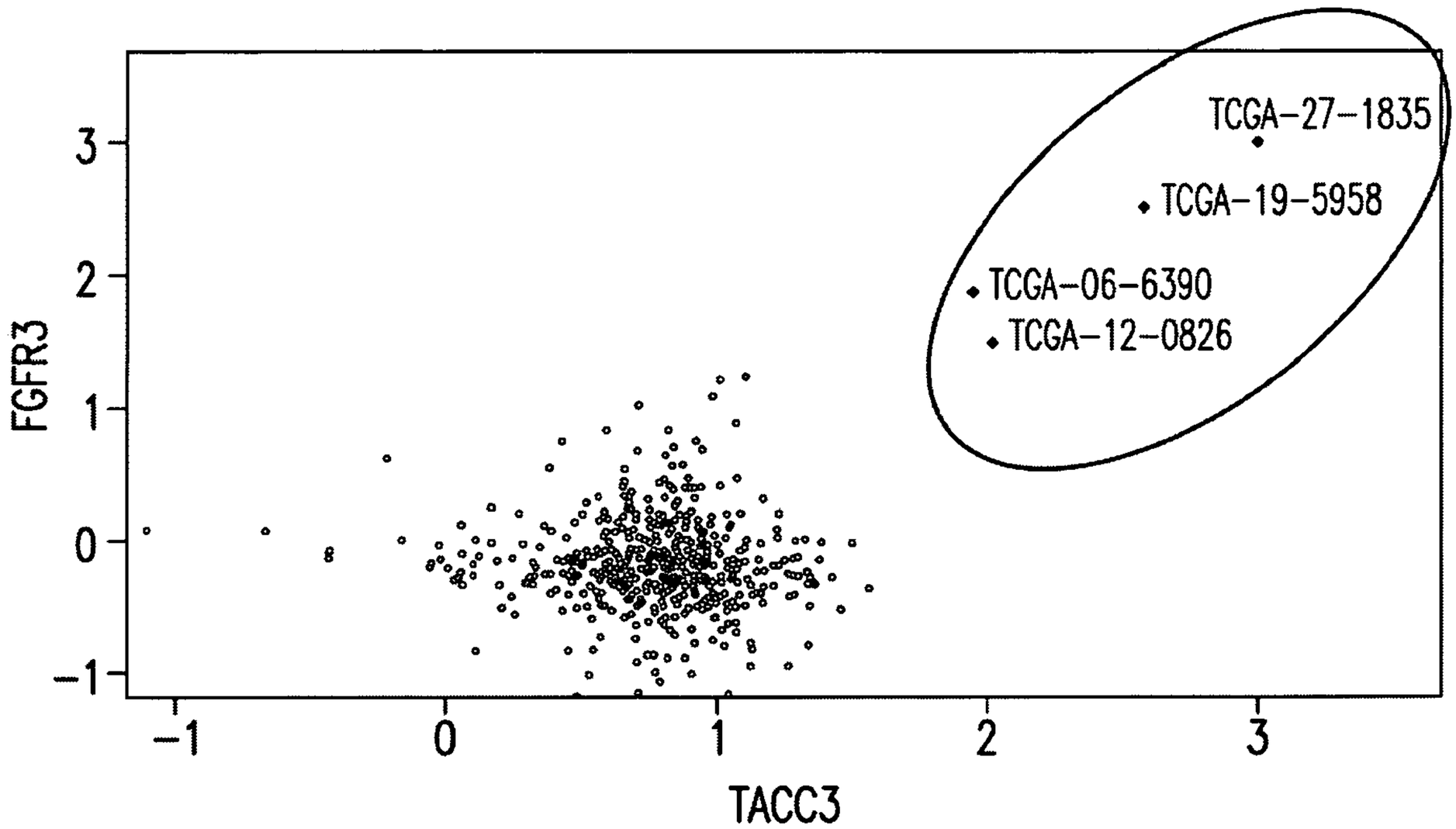

TCGA-27-1835
TCGA-19-5958
TCGA-06-6390
TCGA-12-0826
FGFR3
TACC3
3
2
1
0
−1
−1
0
1
2
3

FIG.2D

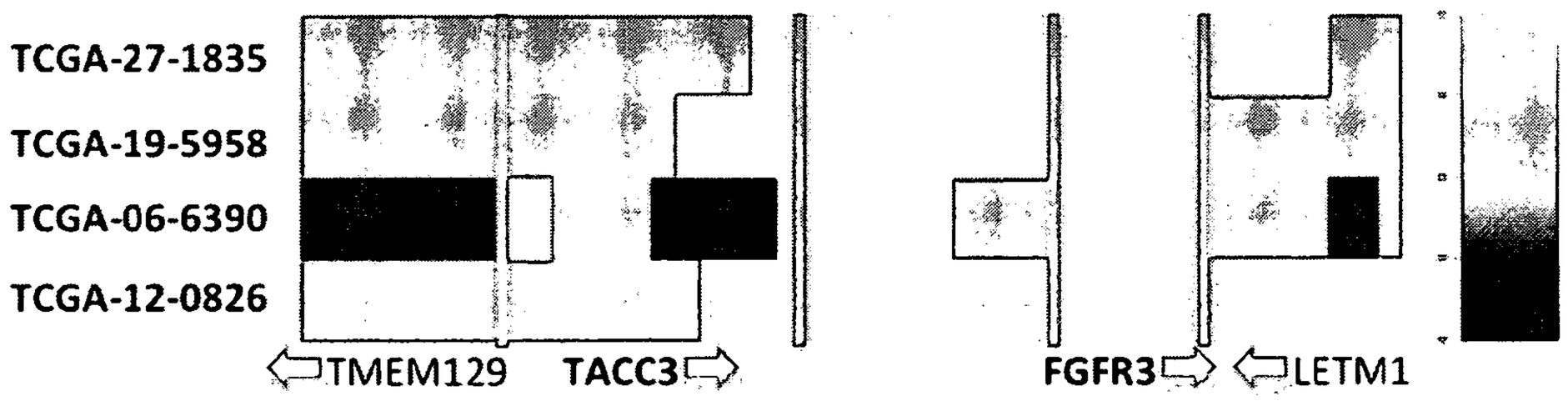

TCGA-27-1835
TCGA-19-5958
TCGA-06-6390
TCGA-12-0826
TMEM129
TACC3
FGFR3
LETM1

FIG.2E

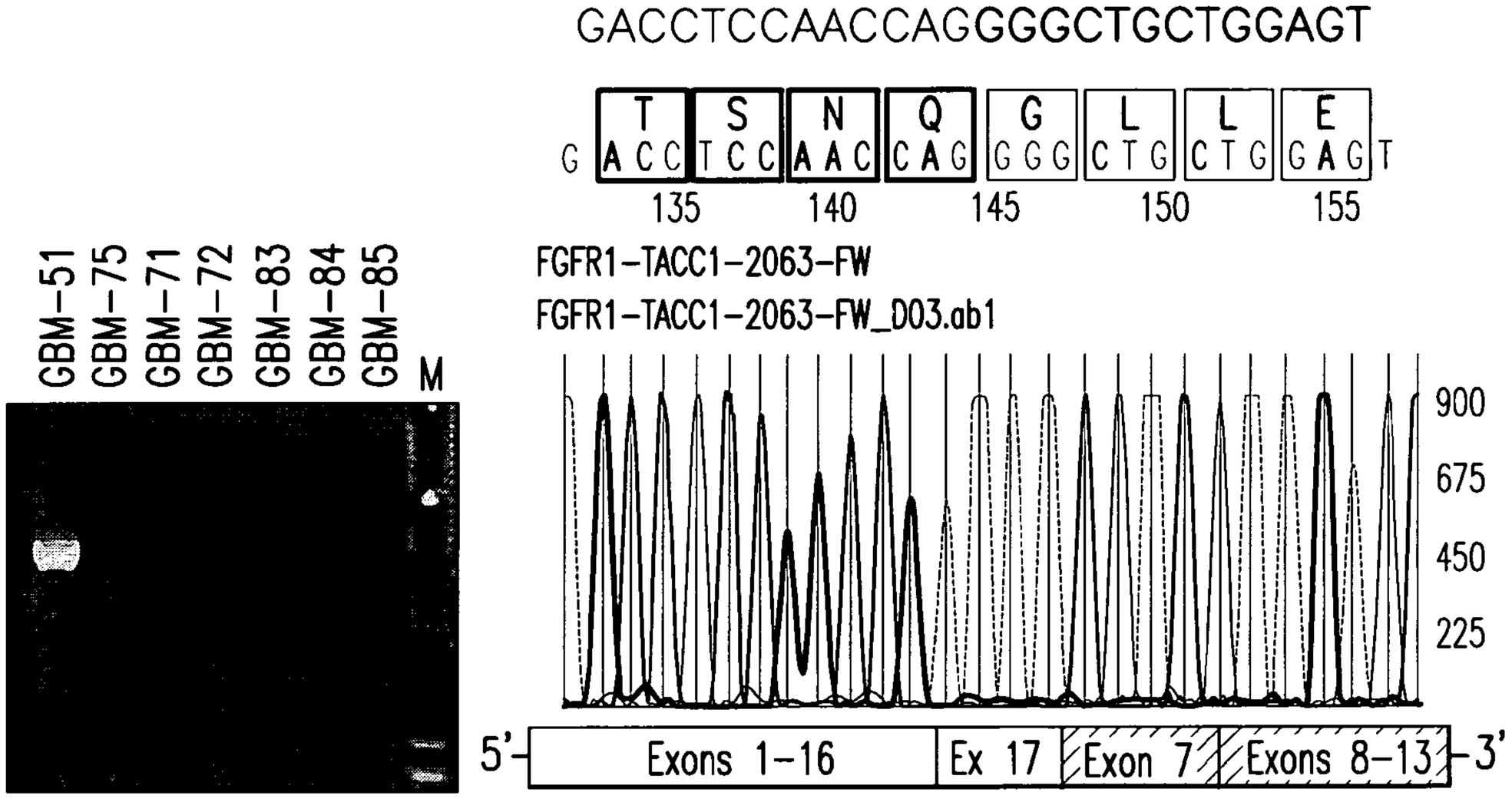

GACCTCCAACCAGGGGCTGCTGGAGT
G T S N Q G L L E
A C C T C C A A C C A G C A G G G G C T G C T G G A G T
135 140 145 150 155
GBM-51
GBM-75
GBM-71
GBM-72
GBM-83
GBM-84
GBM-85
M
FGFR1-TACC1-2063-FW
FGFR1-TACC1-2063-FW_D03.ab1
900
675
450
225
5'
Exons 1-16
Ex 17
Exon 7
Exons 8-13
3'

FIG.2F

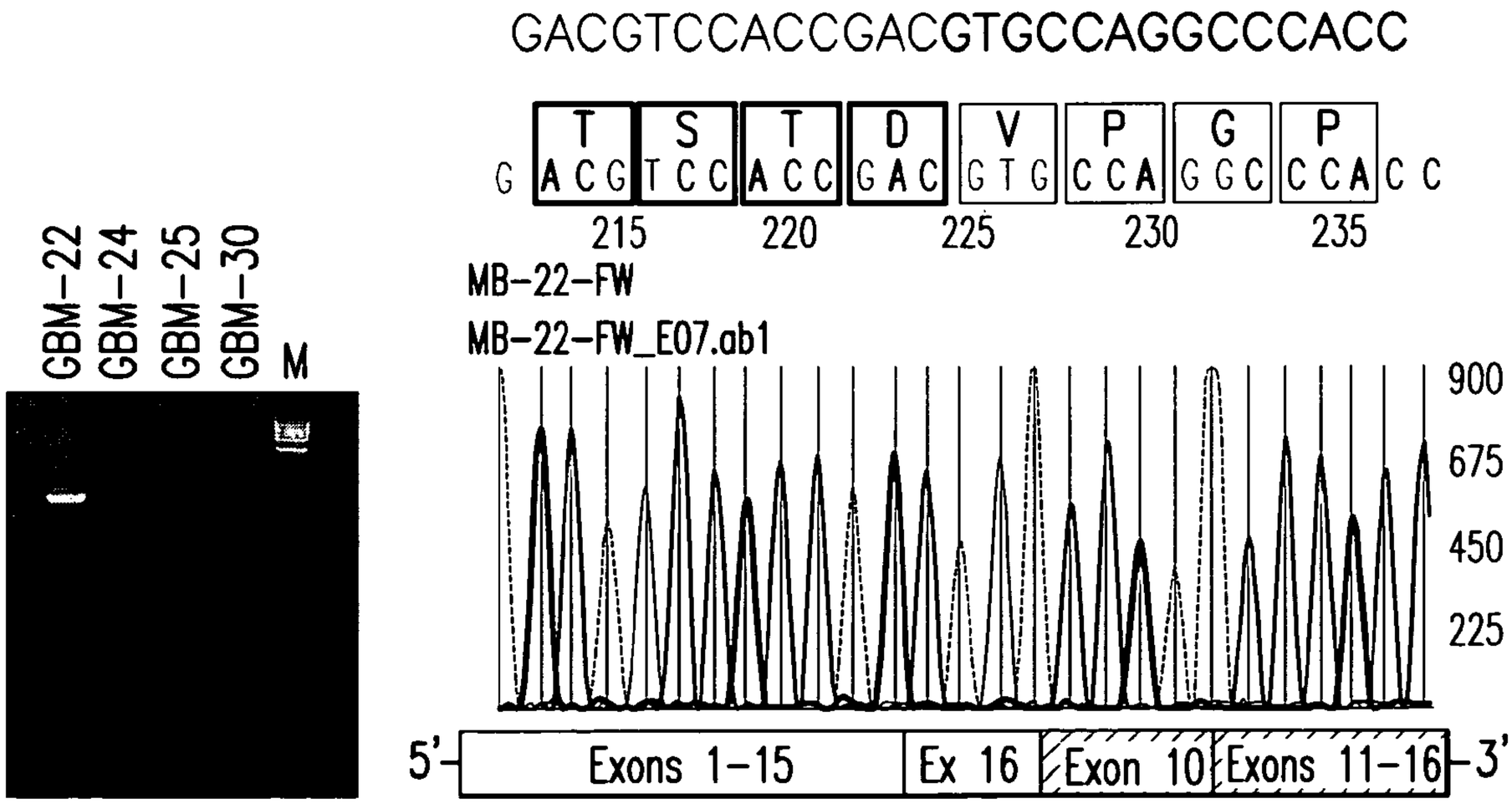

GACGTCCACCGACGTGCCAGGCCCACC
G T S T D V P G P C C
A C G T C C A C C G A C G T G C C A G G C C C A C C
215 220 225 230 235
GBM-22
GBM-24
GBM-25
GBM-30
M
MB-22-FW
MB-22-FW_E07.ab1
900
675
450
225
5'
Exons 1-15
Ex 16
Exon 10
Exons 11-16
3'

FIG.2G

	F1-T1	F3-T3	F3-T3-K508M
Number of colonies	225.3±10.0	198.7±8.0	0

	Vector	FGFR3	TACC3
Number of colonies	0	0	0

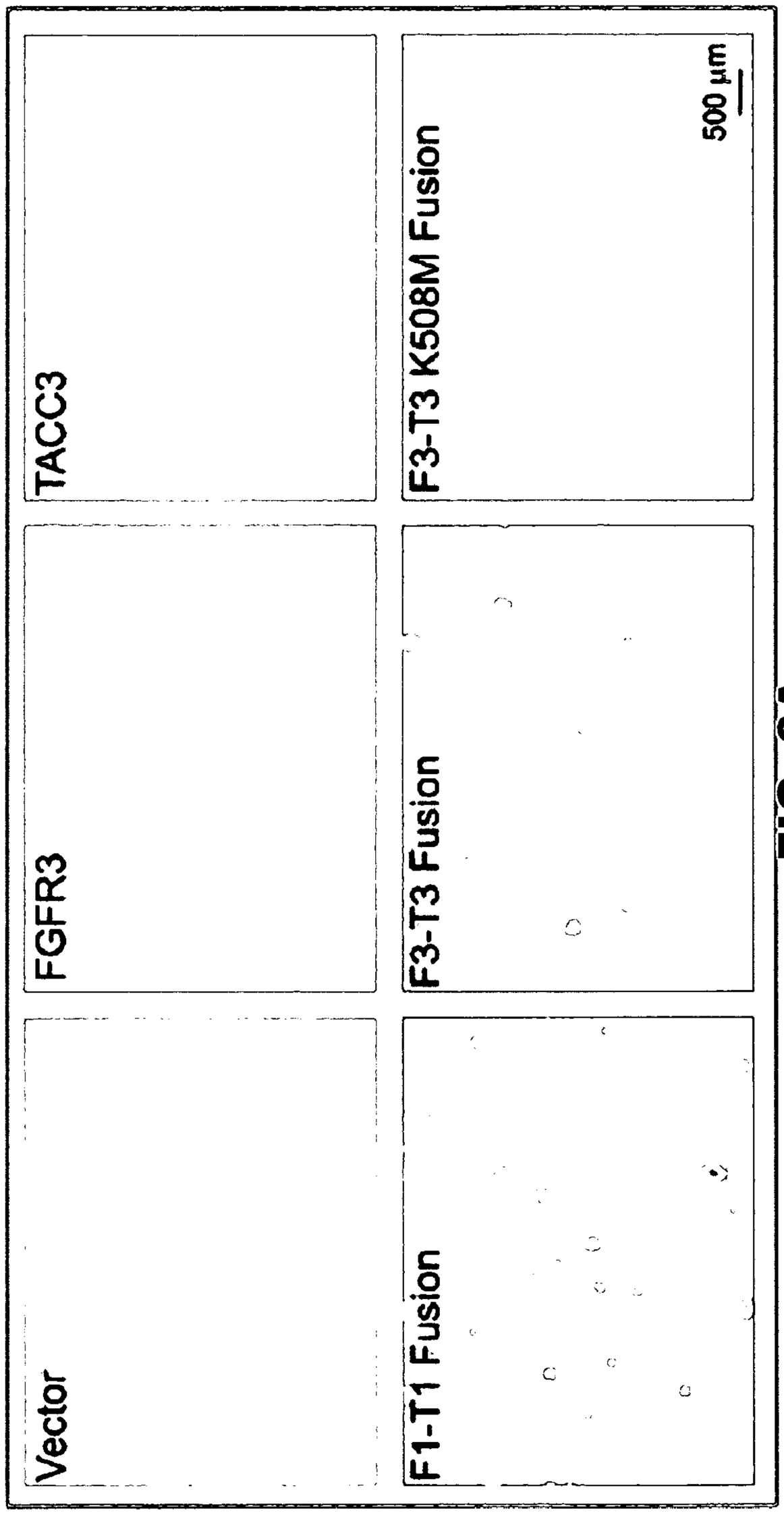


FIG. 3A

	Vector	FGFR3	TACC3	F1-T1	F3-T3	F3-T3-K508M
Number of mice with tumor	0/9	0/5	0/5	8/8	12/12	0/8

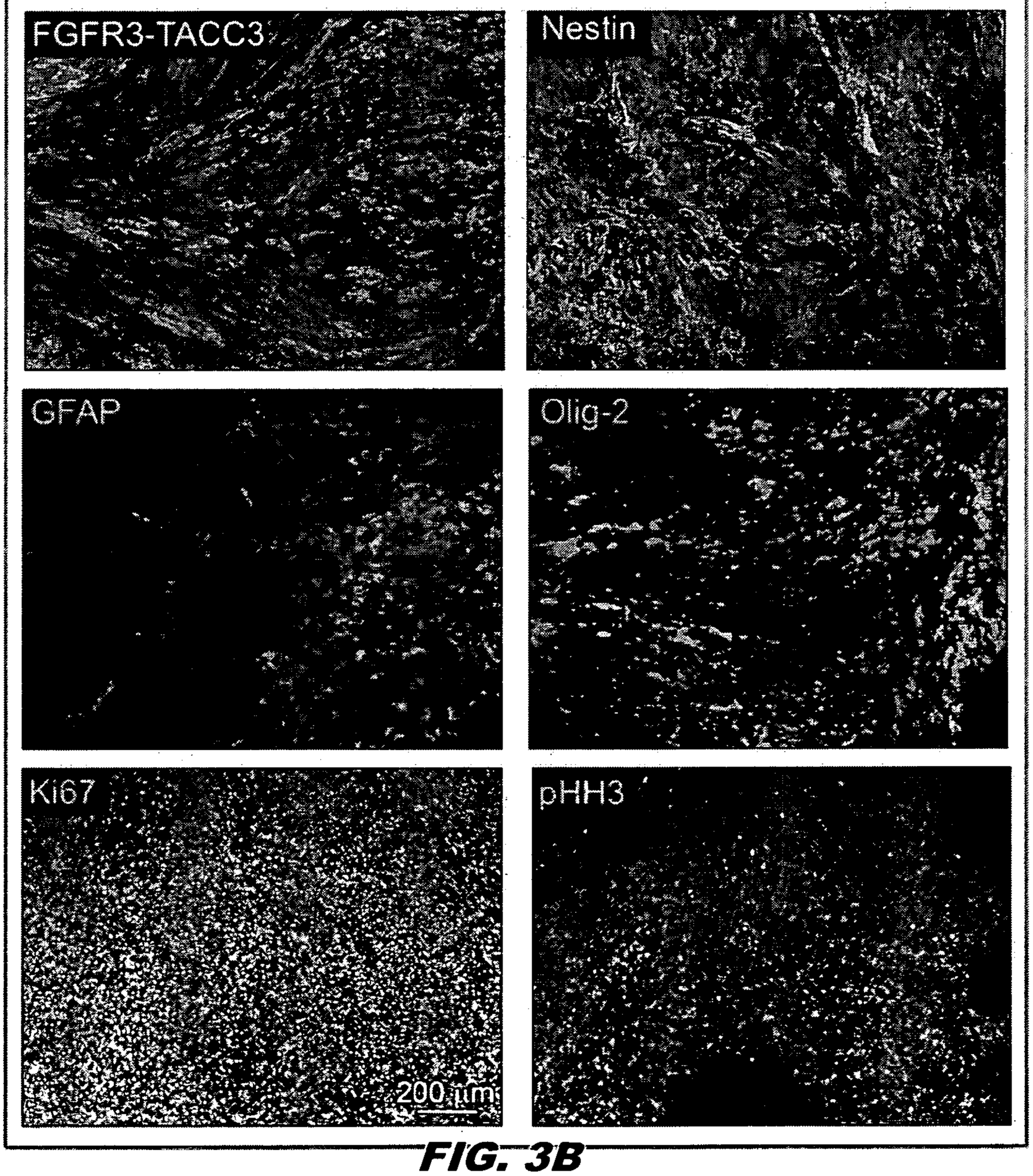


FIG. 3B

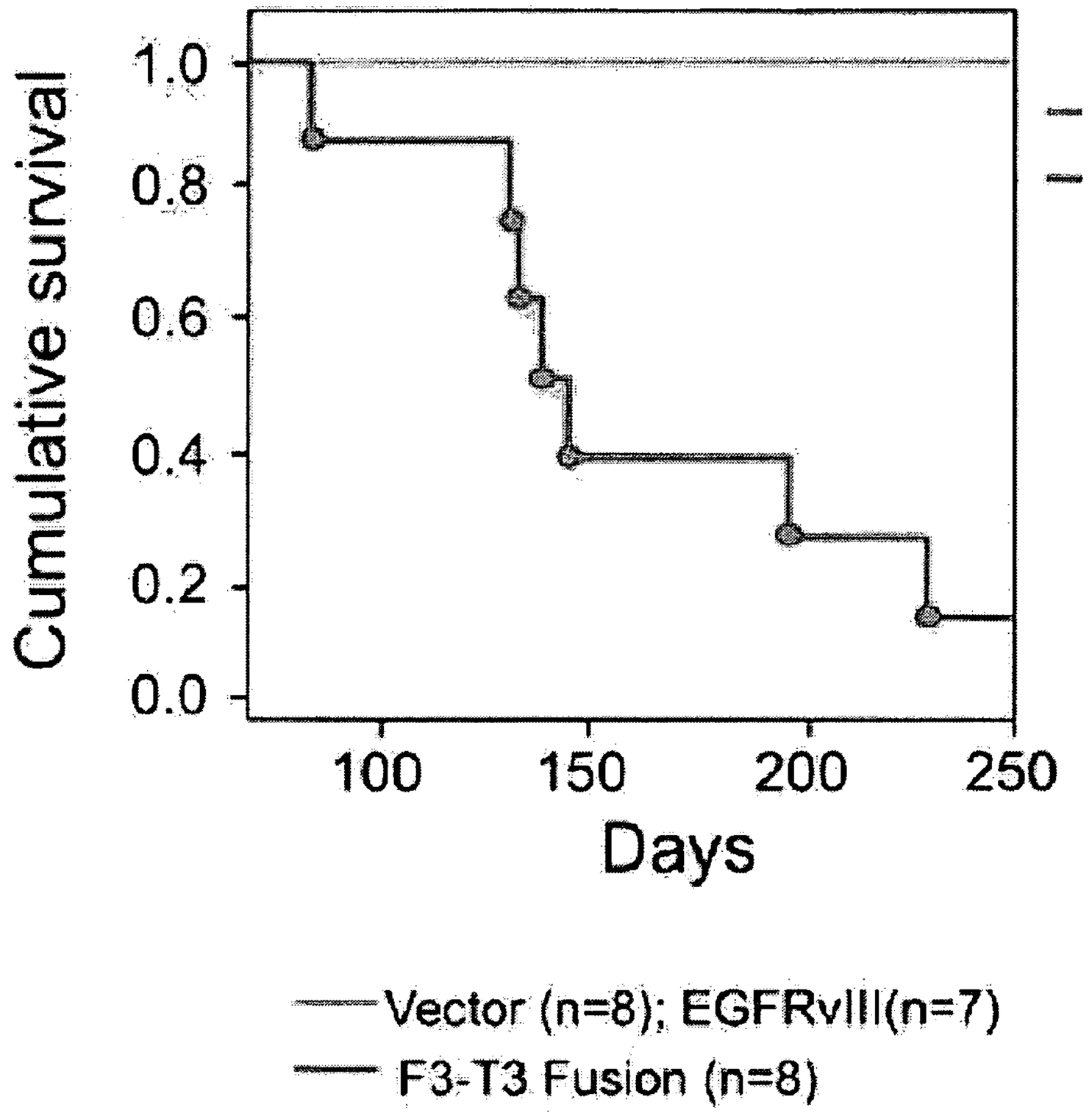

Cumulative survival
1.0
0.8
0.6
0.4
0.2
0.0
100
150
200
250
Days
Vector (n=8); EGFRvIII(n=7)
F3-T3 Fusion (n=8)

FIG. 3C

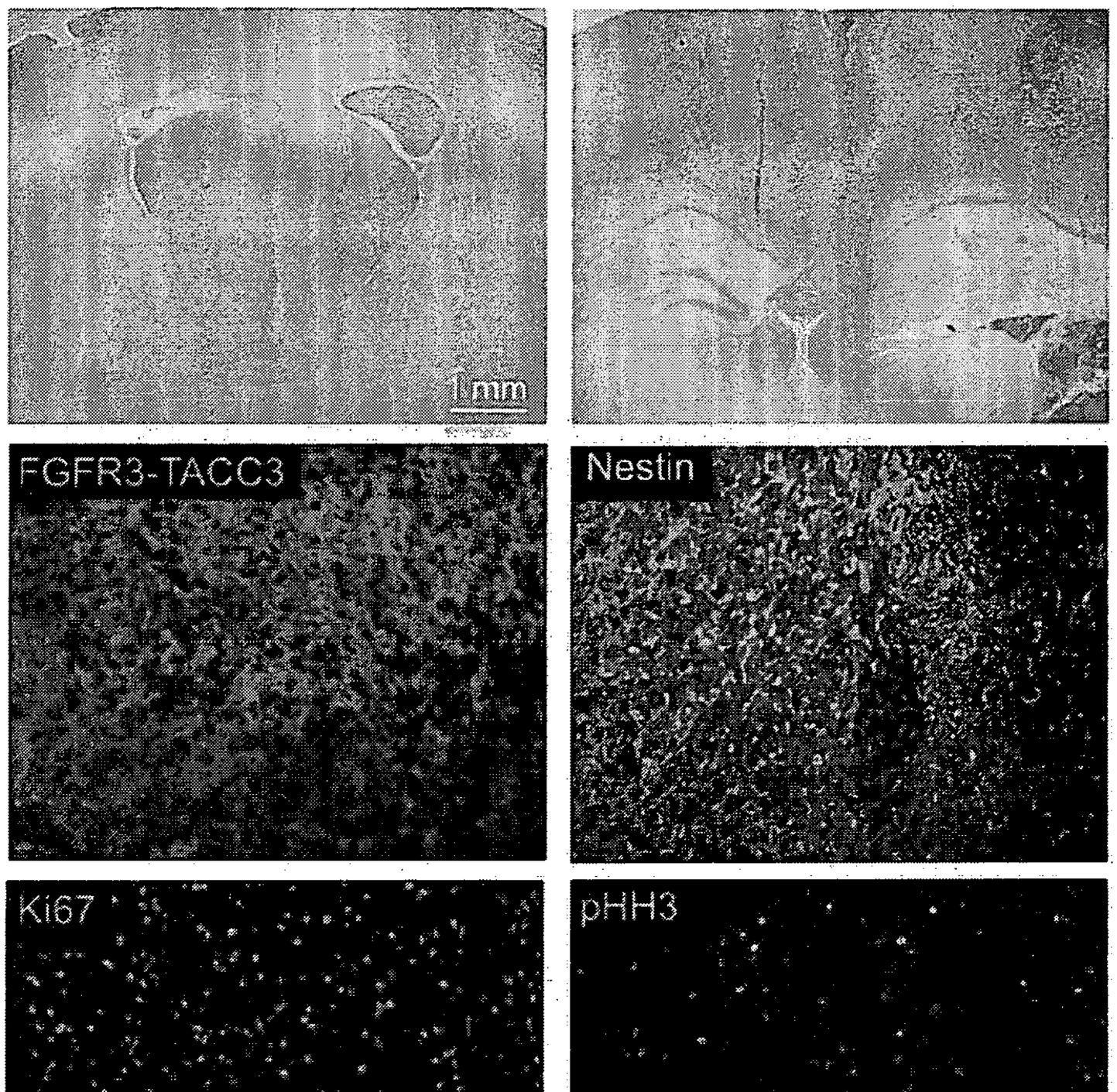


FIG. 3D

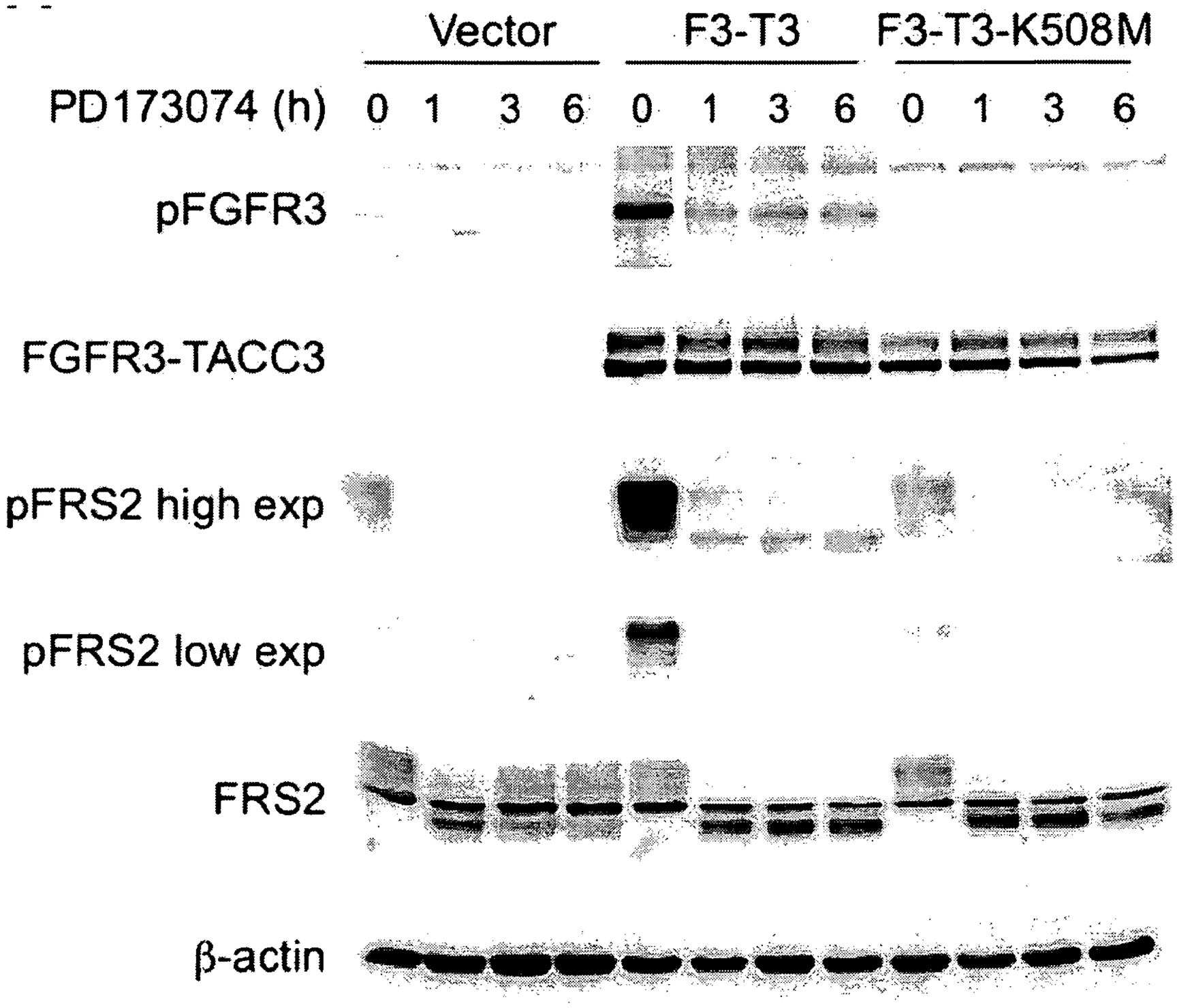

Vector
F3-T3
F3-T3-K508M
PD173074 (h)
0 1 3 6 0 1 3 6 0 1 3 6
pFGFR3
FGFR3-TACC3
pFRS2 high exp
pFRS2 low exp
FRS2
β-actin

FIG. 4A

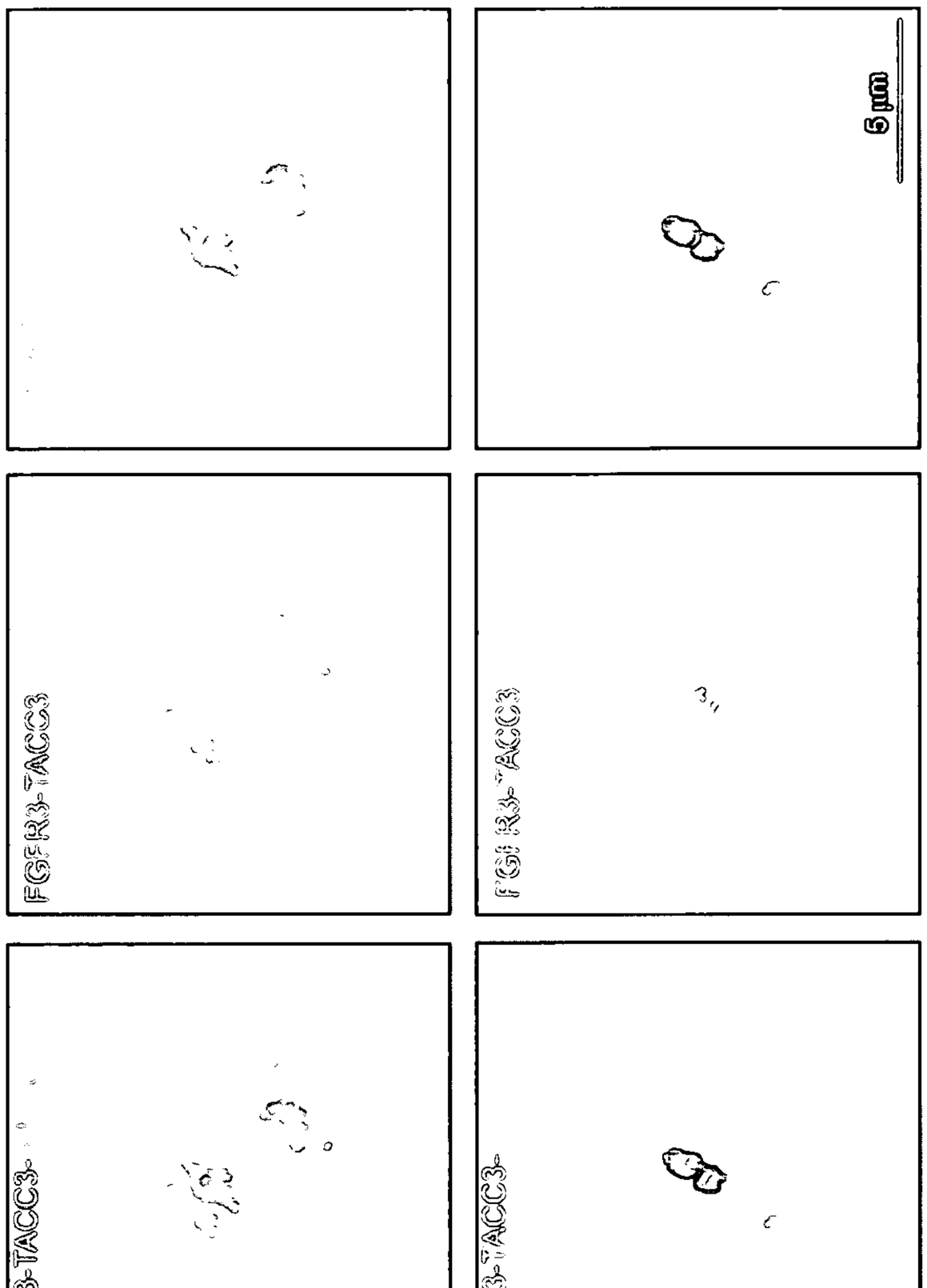

FGFR3-TACC3
5 µm
FGFR3-TACC3
FGFR3-TACC3
FGFR3-TACC3
FGFR3-TACC3
FGFR3-TACC3

FIG. 48

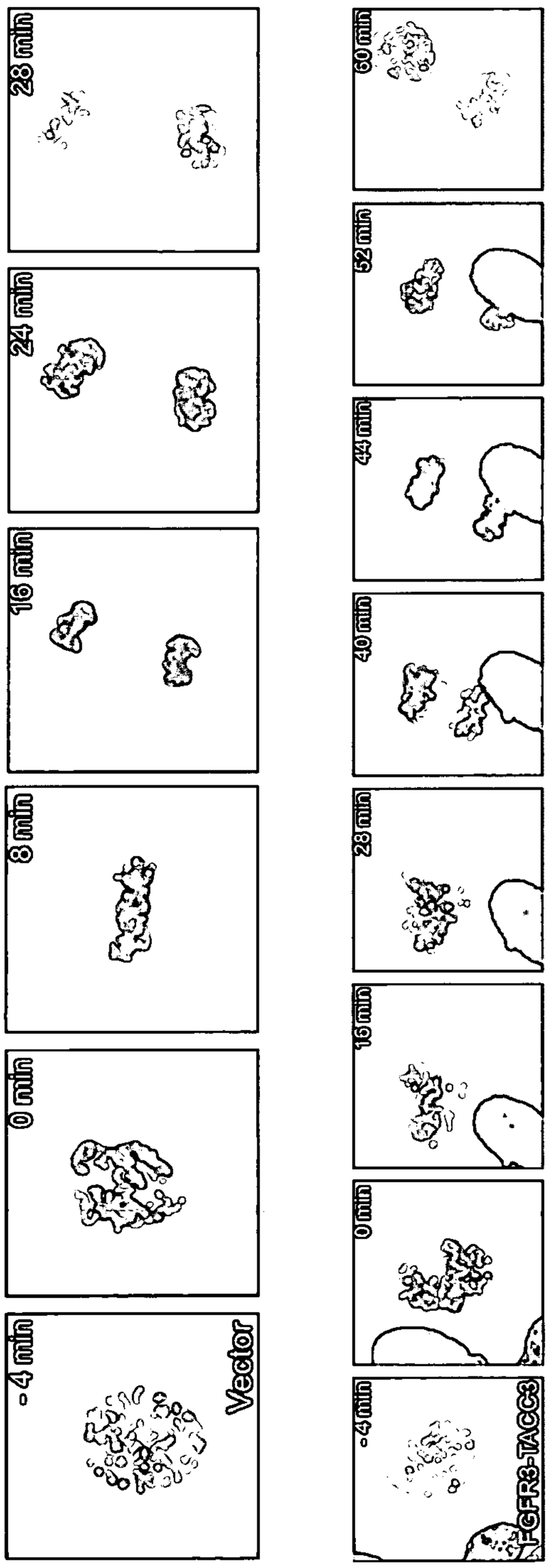

Vector
0 min
8 min
16 min
24 min
28 min
40 min
44 min
52 min
60 min
FGFR3-TACC3
0 min
16 min
28 min
40 min
44 min
52 min
60 min

FIG. 4C

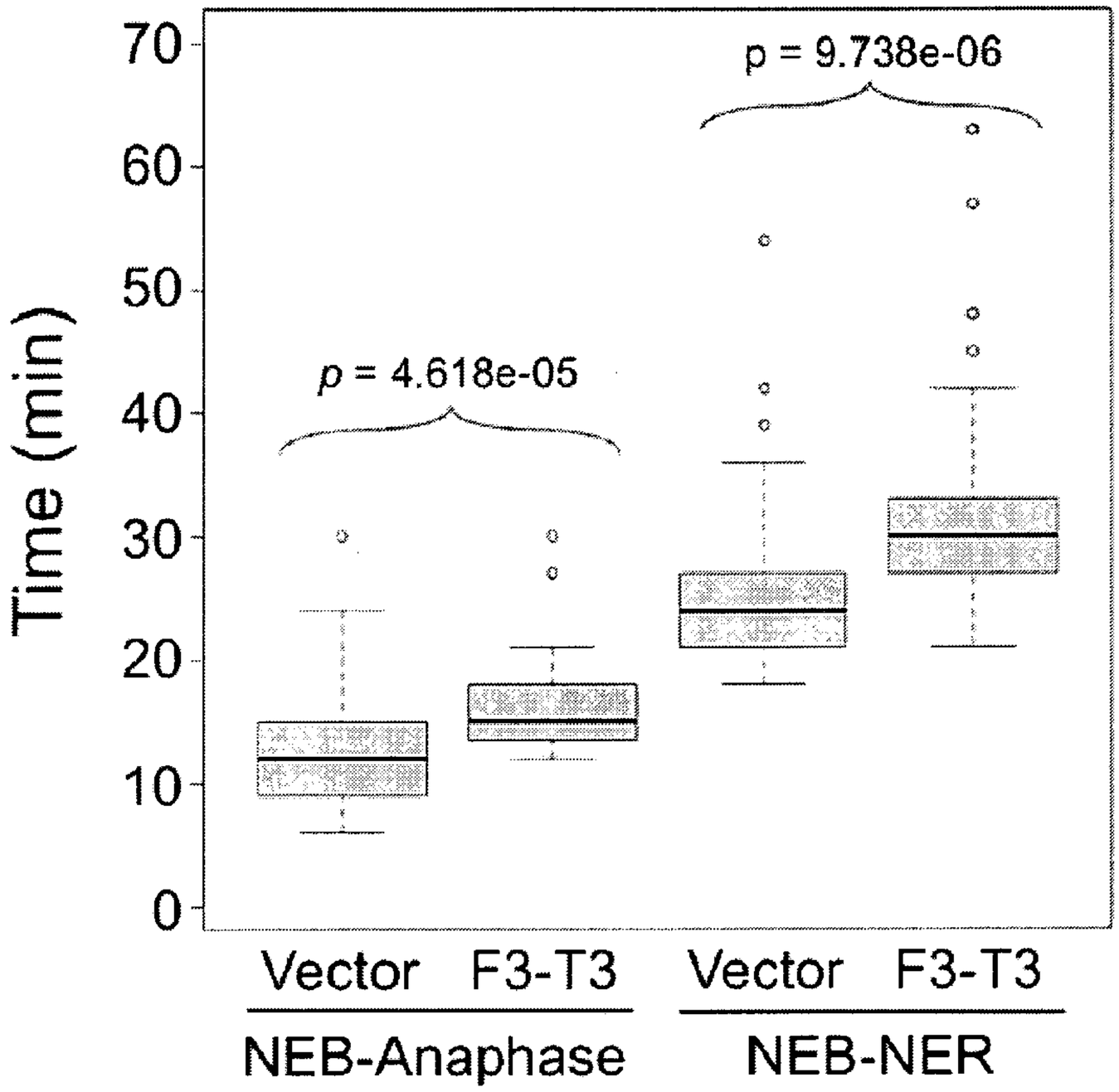

70
60
50
40
30
20
10
0
Time (min)
p = 9.738e-06
p = 4.618e-05
Vector
F3-T3
Vector
F3-T3
NEB-Anaphase
NEB-NER

FIG. 4D

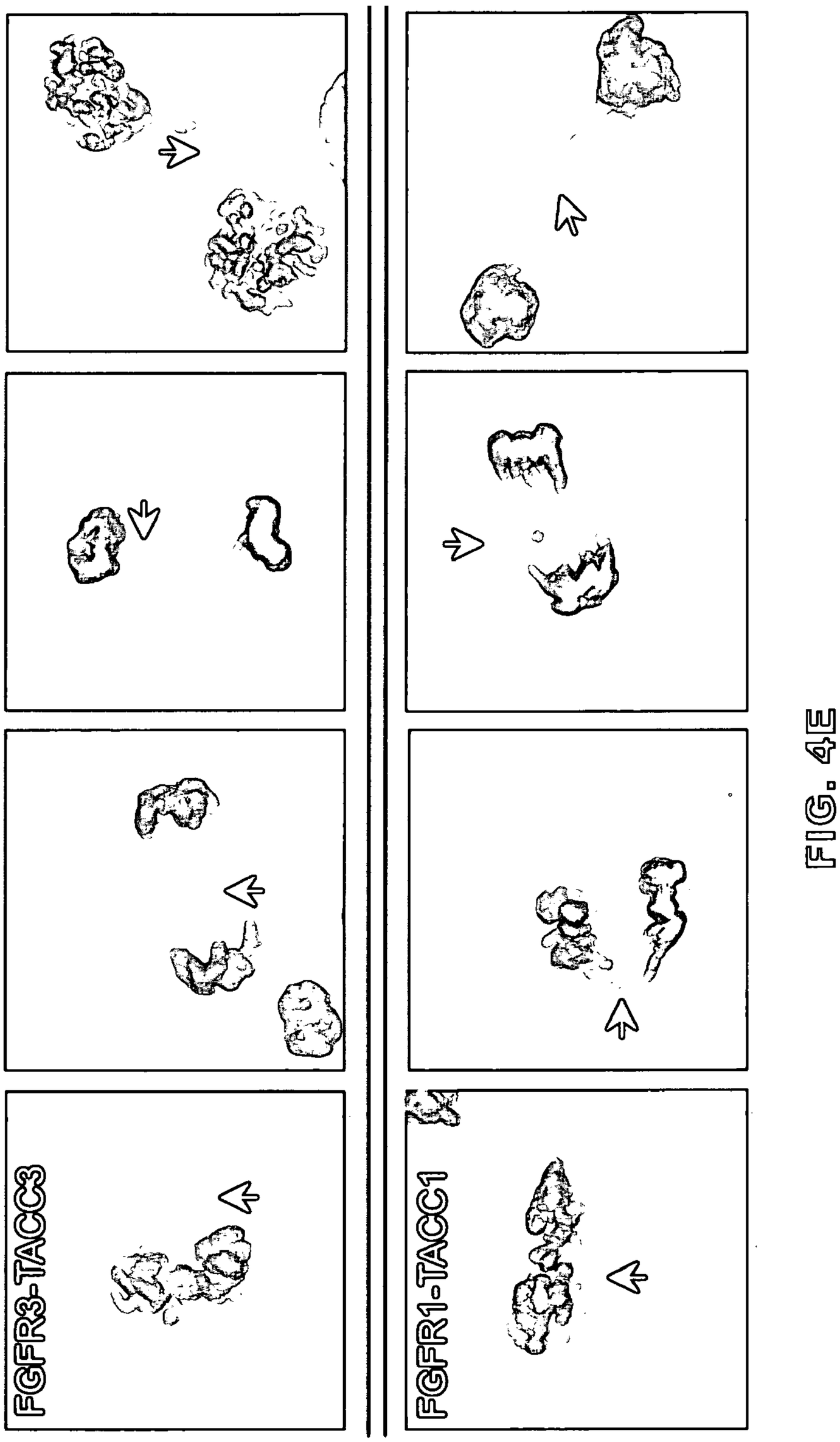


FIG. 4E

Cell line	Metaphases inspected	Cells with segregation defects (%±SD)	Metaphases with misaligned chromosomes (%±SD)	Anaphases with lagging chromosomes (%±SD)	Anaphases with bridges (%±SD)
Rat1A Vector	150	9.5±3.8	2.3±1.7	2.4±0.9	5.4±2.4
Rat1A FGFR3–TACC3	150	27.4±3.9	8.0±2.6	8.0±1.0	11.2±0.7
Rat1A FGFR1–TACC1	100	45.5±3.1	5±2.8	10±1.4	31.5±4.9

FIG.4F

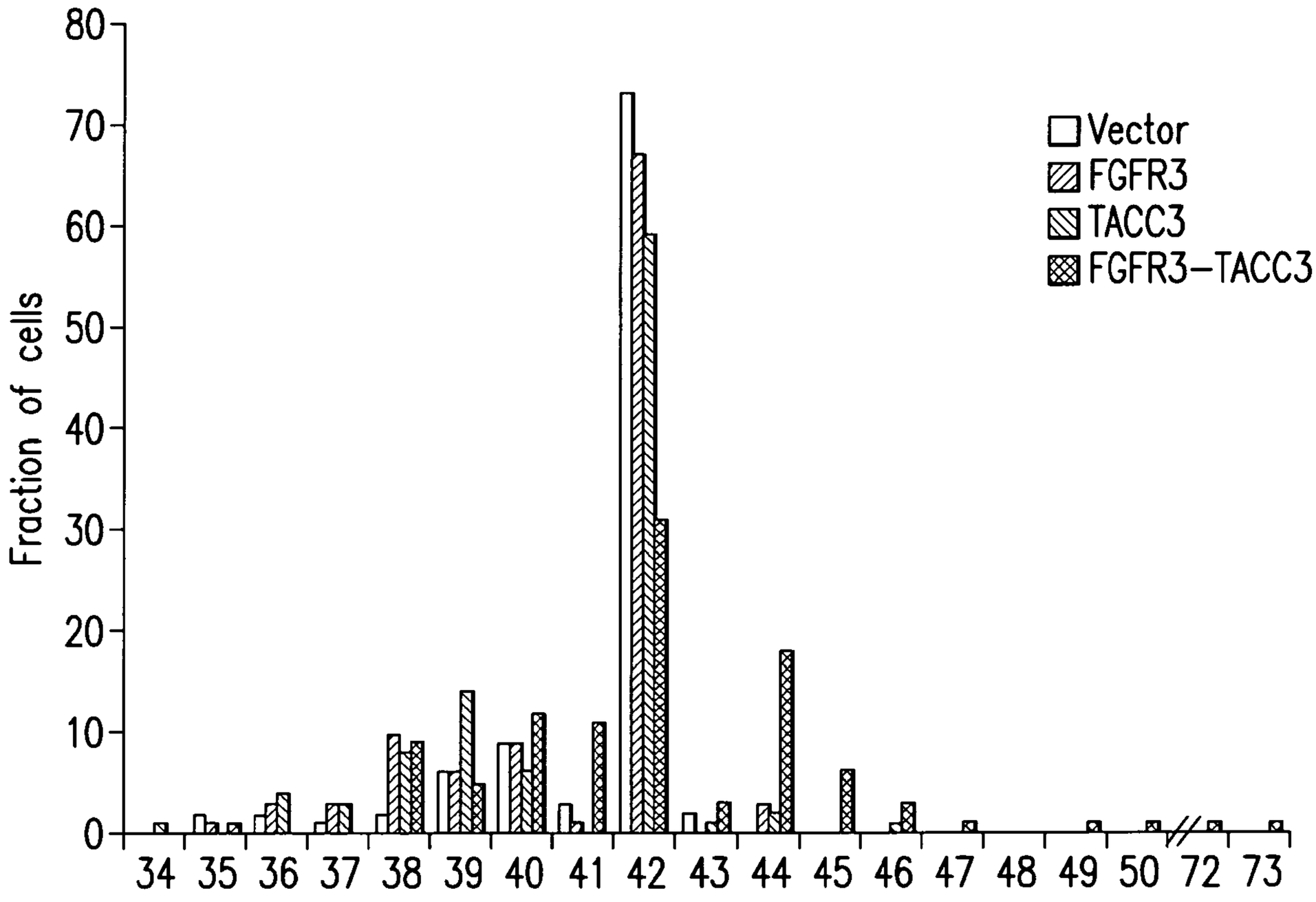

Fraction of cells
80
70
60
50
40
30
20
10
0
34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 72 73
Vector
FGFR3
TACC3
FGFR3–TACC3

FIG.5A

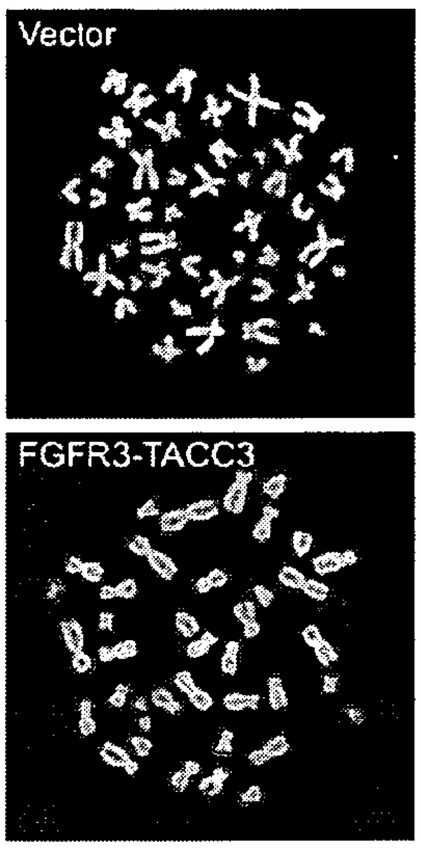

Vector
FGFR3-TACC3

FIG. 5B

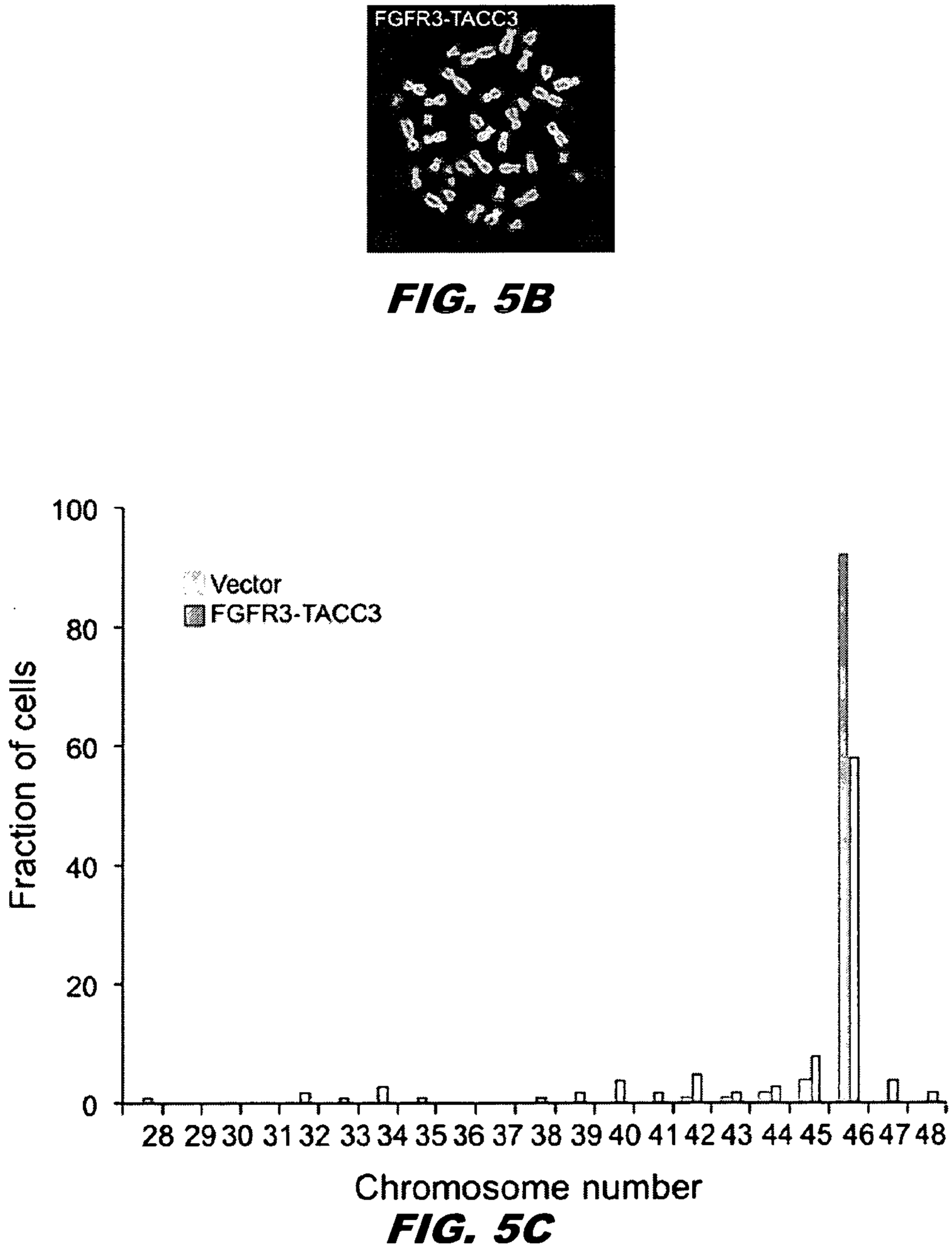
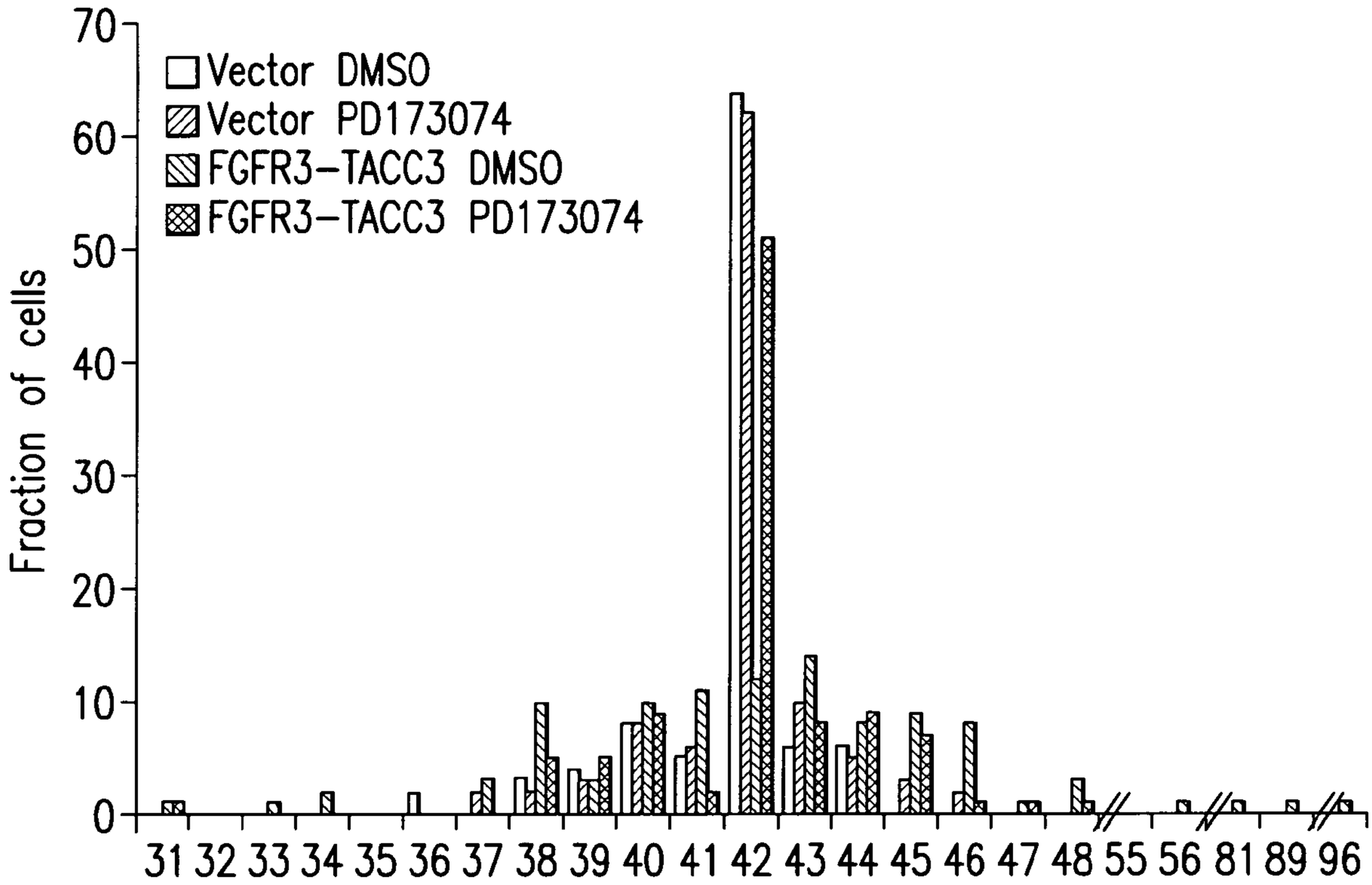

Vector
FGFR3-TACC3
Fraction of cells
100
80
60
40
20
0
28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48
Chromosome number

FIG. 5C

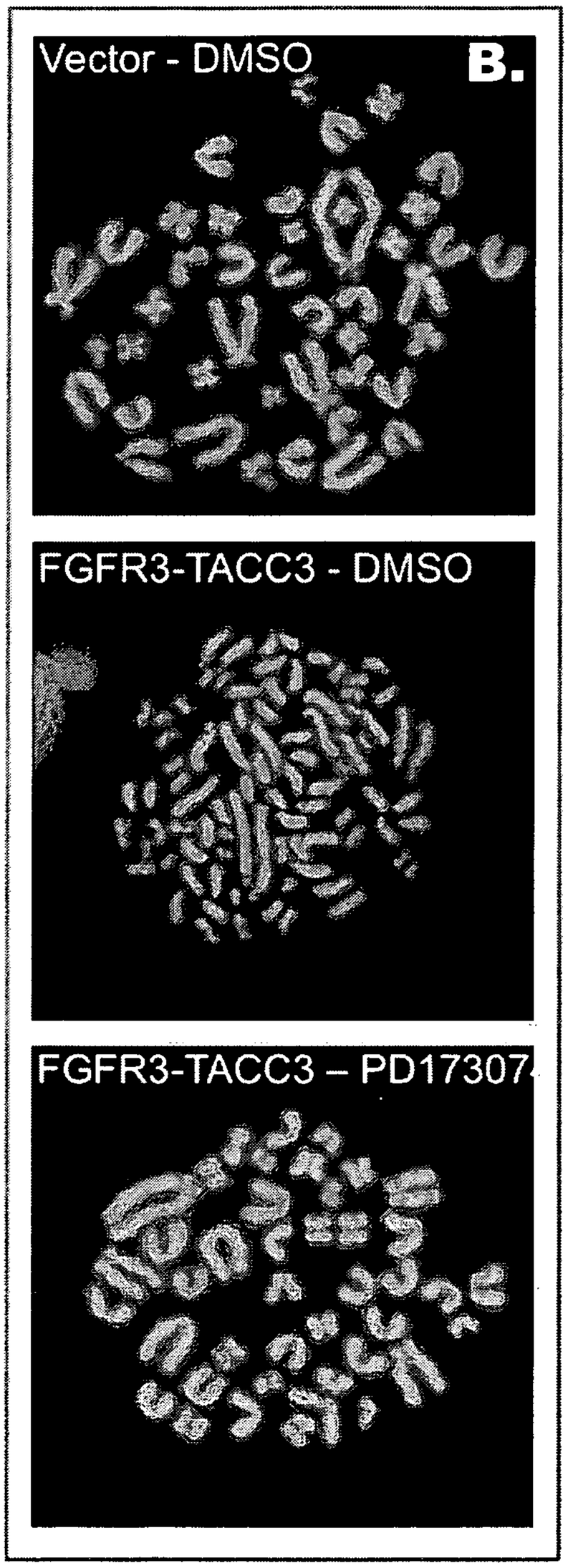
Cell line	Number of cells counted	Percent aneuploidy	Range	Mean number	Average variation from mean number	*p*- value
Vector	100	8	42-46	45.85	0.28	$p = <0.001$
FGFR3-TACC3	100	42	28-48	42.24	3.33	

FIG. 5D



Cell line	Number of cells counted	Percent aneuploidy	Range	Mean number	Average variation from mean number	*p*-value
Rat1A Vector DMSO	100	38	37–55	41.7	1.10	
Rat1A Vector PD173074	100	39	37–46	42.0	0.77	n.s.
Rat1A FGFR3-TACC3 DMSO	100	88	31–96	43.7	4.11	<0.0001
Rat1A FGFR3-TACC3 PD173074	100	49	31–48	42.0	1.28	<0.0001

FIG.6A


Vector - DMSO
B.
FGFR3-TACC3 - DMSO
FGFR3-TACC3 – PD17307
C.
Vector DMSO
Vector PD173074
FGFR3-TACC3 DMSO
FGFR3-TACC3 PD173074
Fraction of PMSCS
20
15
10
5
0
*

FIGS. 6B-C

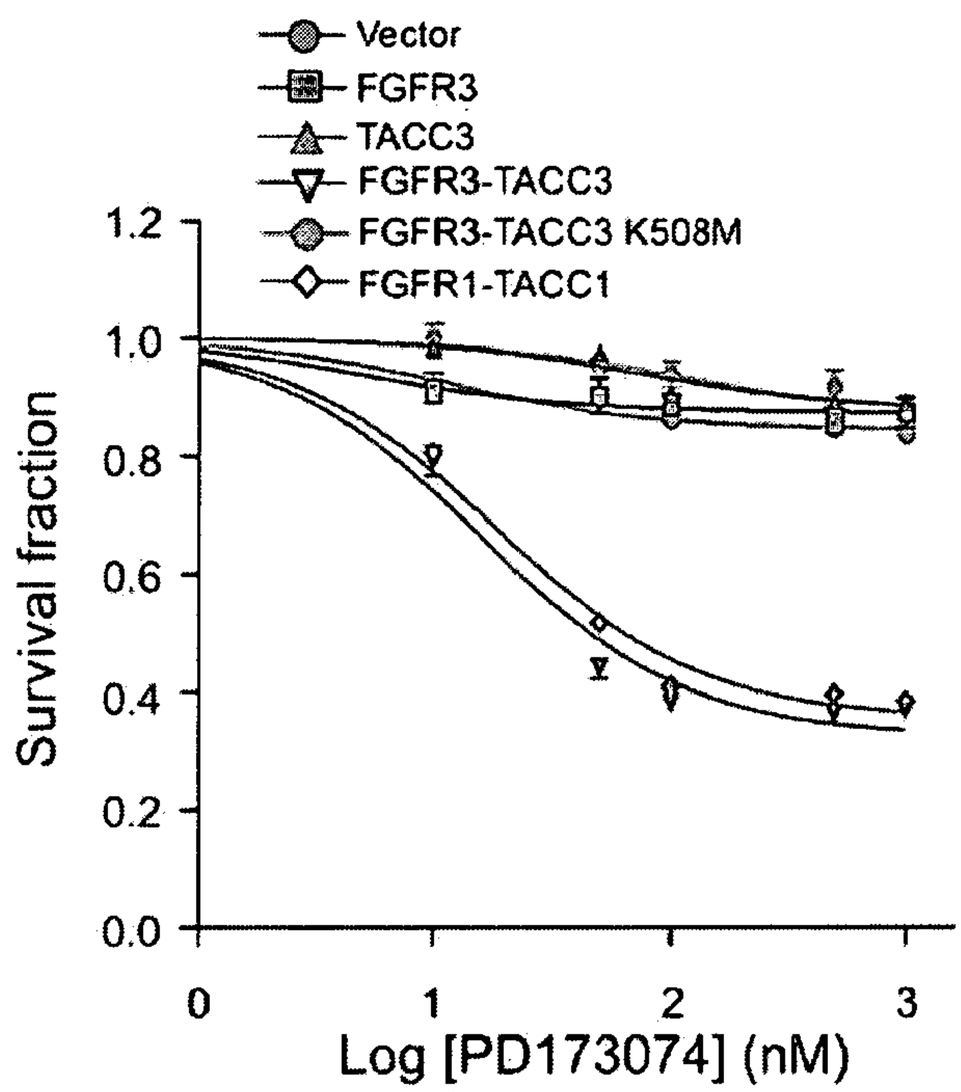

Vector
FGFR3
TACC3
FGFR3-TACC3
FGFR3-TACC3 K508M
FGFR1-TACC1
Survival fraction
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
Log [PD173074] (nM)

FIG. 7A

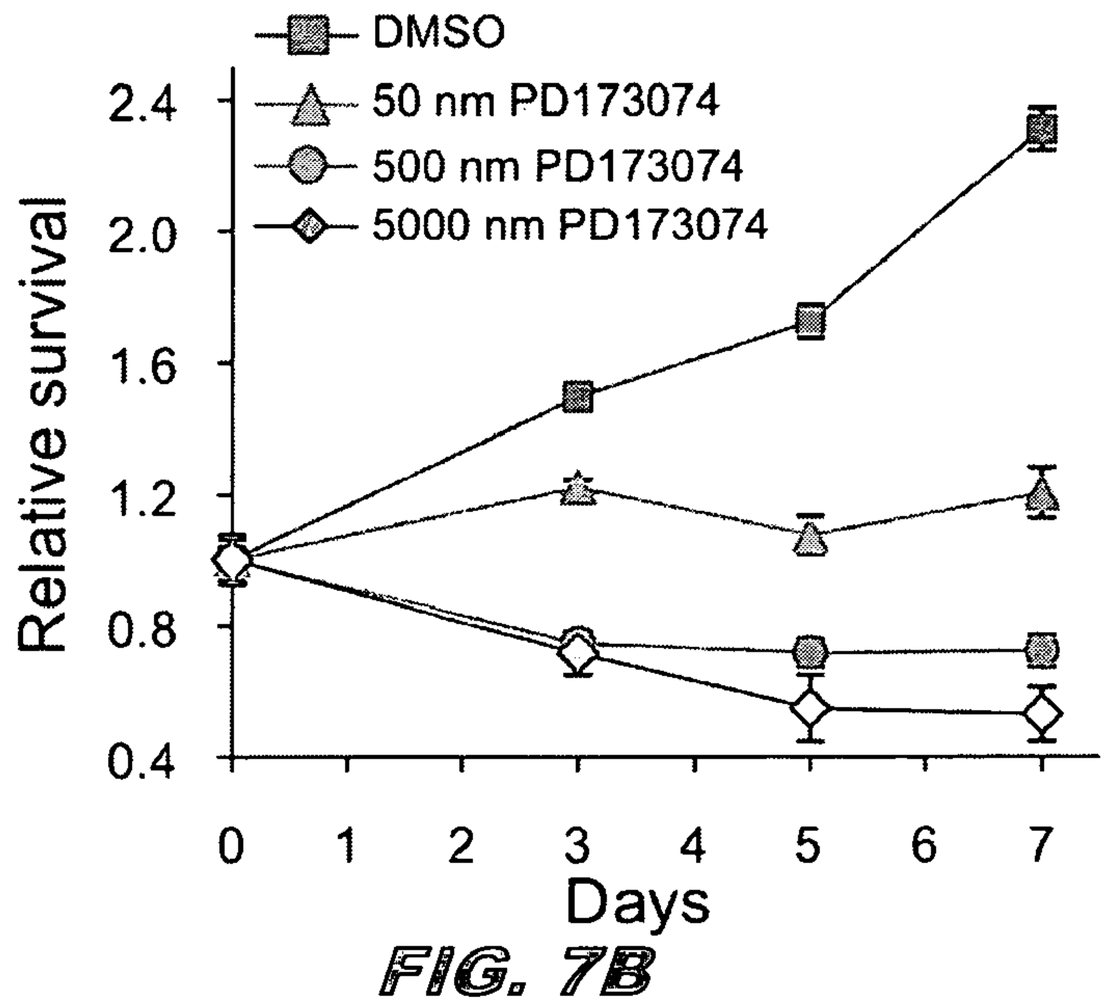

DMSO
50 nm PD173074
500 nm PD173074
5000 nm PD173074
Relative survival
2.4
2.0
1.6
1.2
0.8
0.4
0
1
2
3
4
5
6
7
Days

FIG. 7B

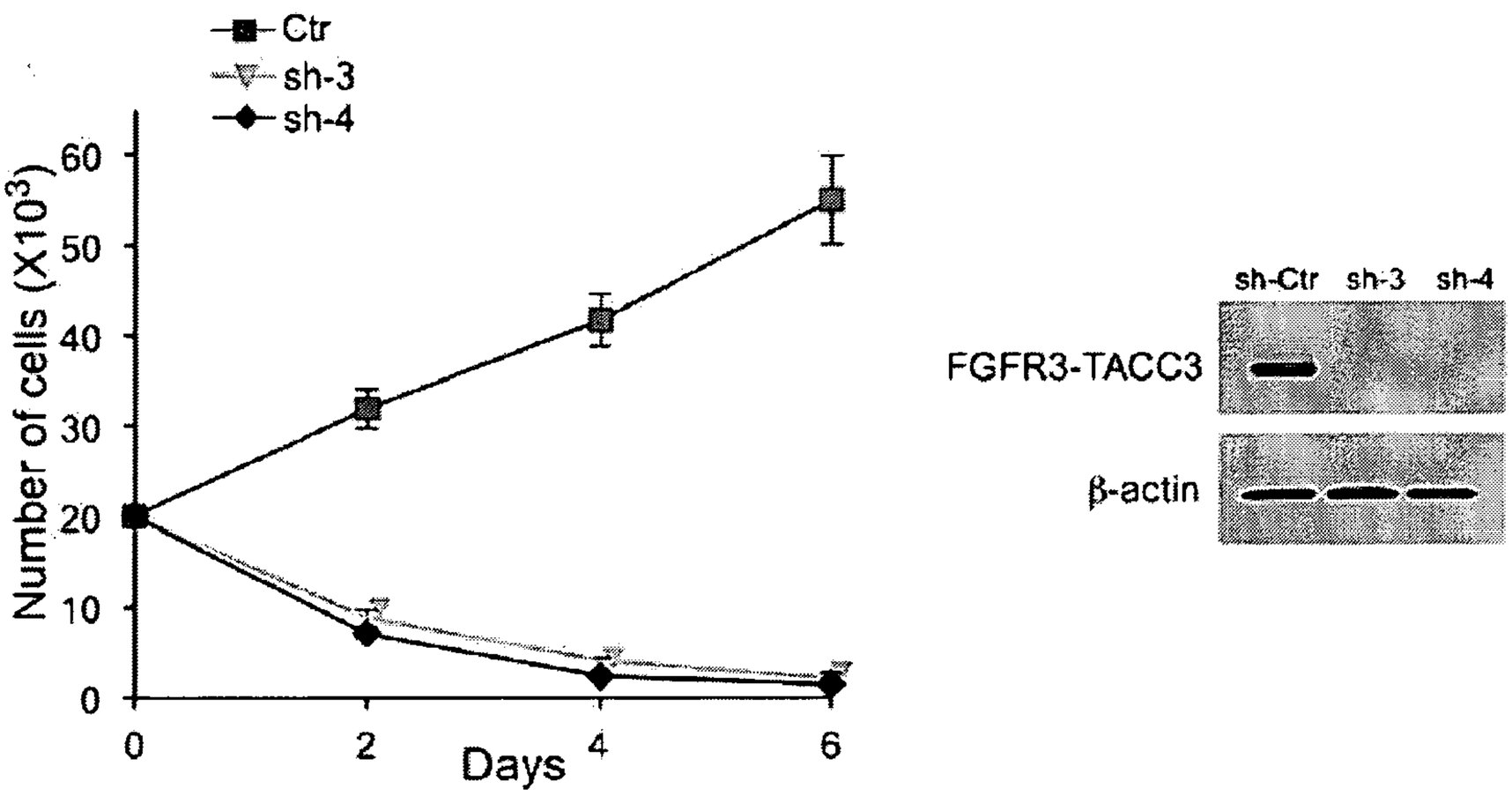

Ctr
sh-3
sh-4
Number of cells (X10^3)
60
50
40
30
20
10
0
0
2
4
6
Days
sh-Ctr
sh-3
sh-4
FGFR3-TACC3
β-actin

FIG. 7C

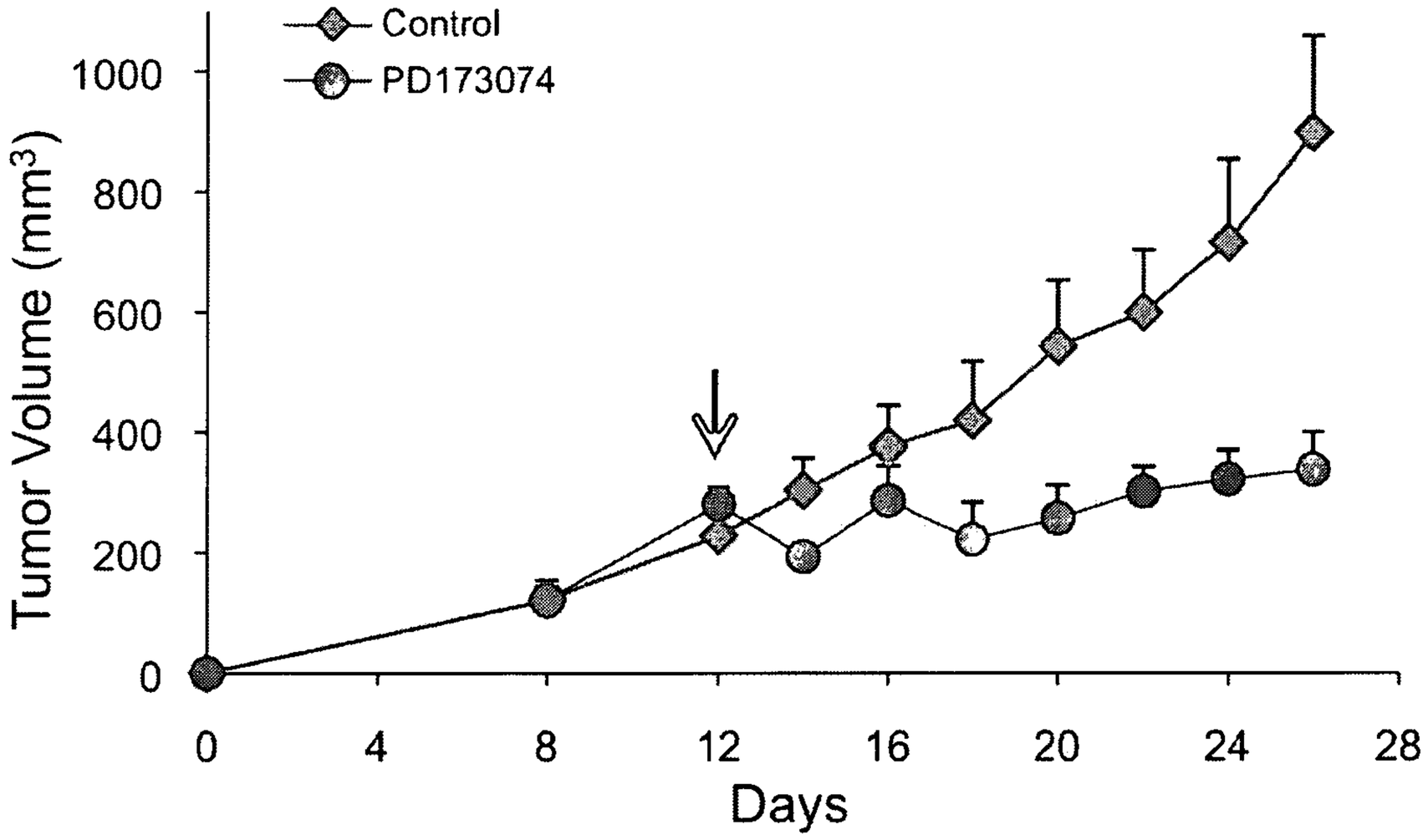

Control
PD173074
Tumor Volume (mm3)
1000
800
600
400
200
0
0
4
8
12
16
20
24
28
Days

FIG. 7D

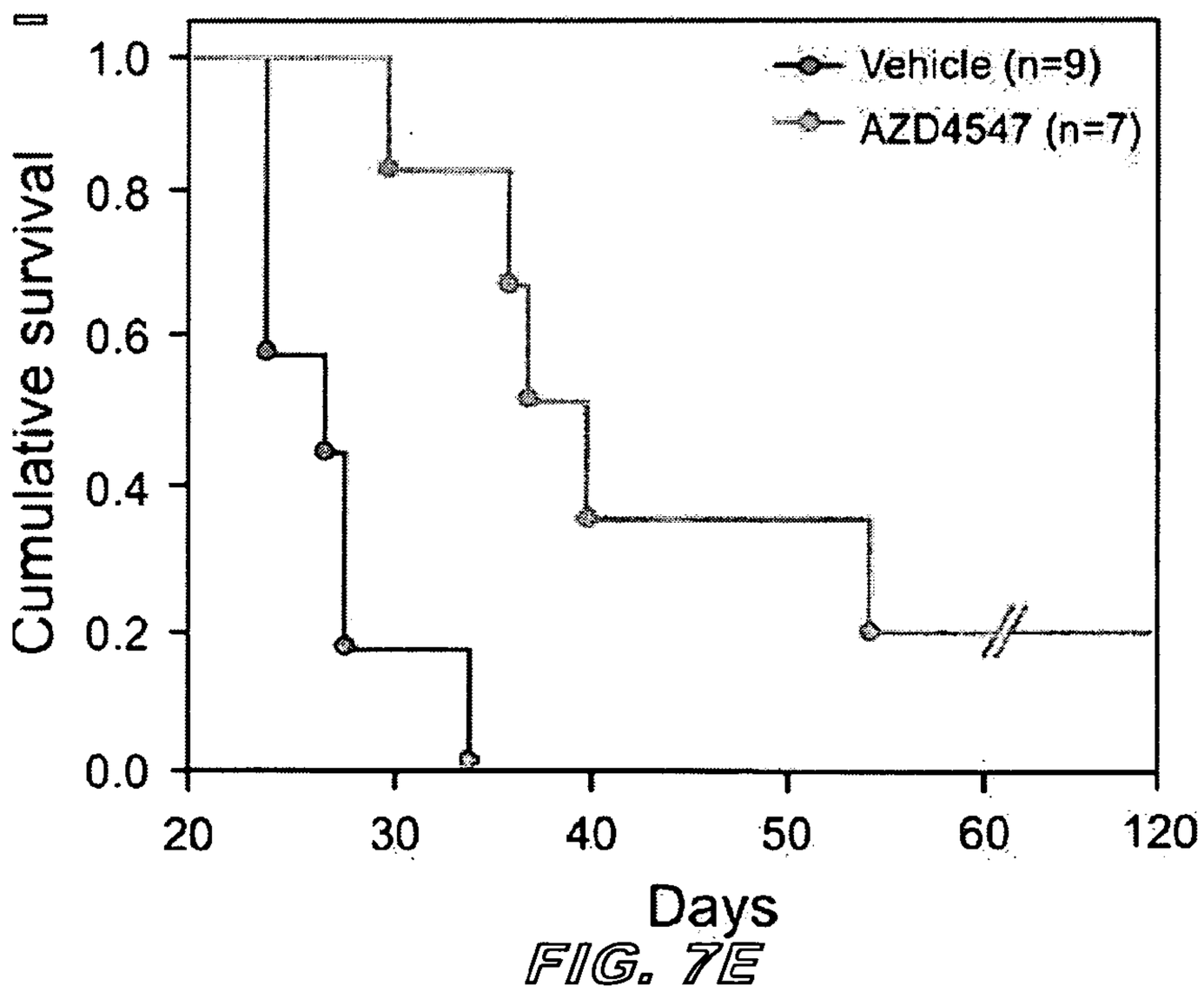

Vehicle (n=9)
AZD4547 (n=7)
Cumulative survival
1.0
0.8
0.6
0.4
0.2
0.0
20
30
40
50
60
120
Days

FIG. 7E

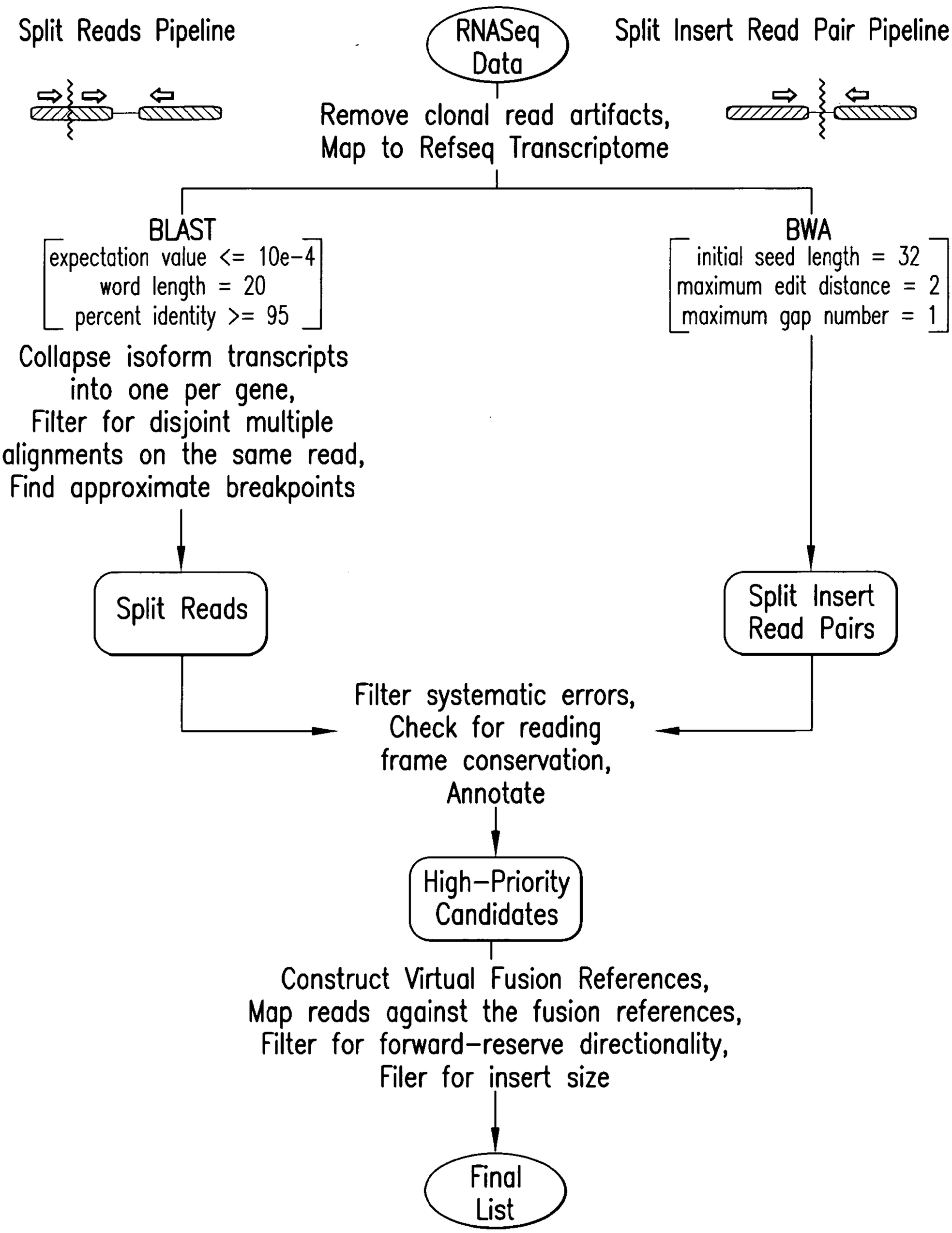

Split Reads Pipeline
RNASeq Data
Split Insert Read Pair Pipeline
Remove clonal read artifacts,
Map to Refseq Transcriptome
BLAST
expectation value <= 10e-4
word length = 20
percent identity >= 95
BWA
initial seed length = 32
maximum edit distance = 2
maximum gap number = 1
Collapse isoform transcripts into one per gene,
Filter for disjoint multiple alignments on the same read,
Find approximate breakpoints
Split Reads
Split Insert Read Pairs
Filter systematic errors,
Check for reading frame conservation,
Annotate
High-Priority Candidates
Construct Virtual Fusion References,
Map reads against the fusion references,
Filter for forward-reserve directionality,
Filer for insert size
Final List

FIG.8

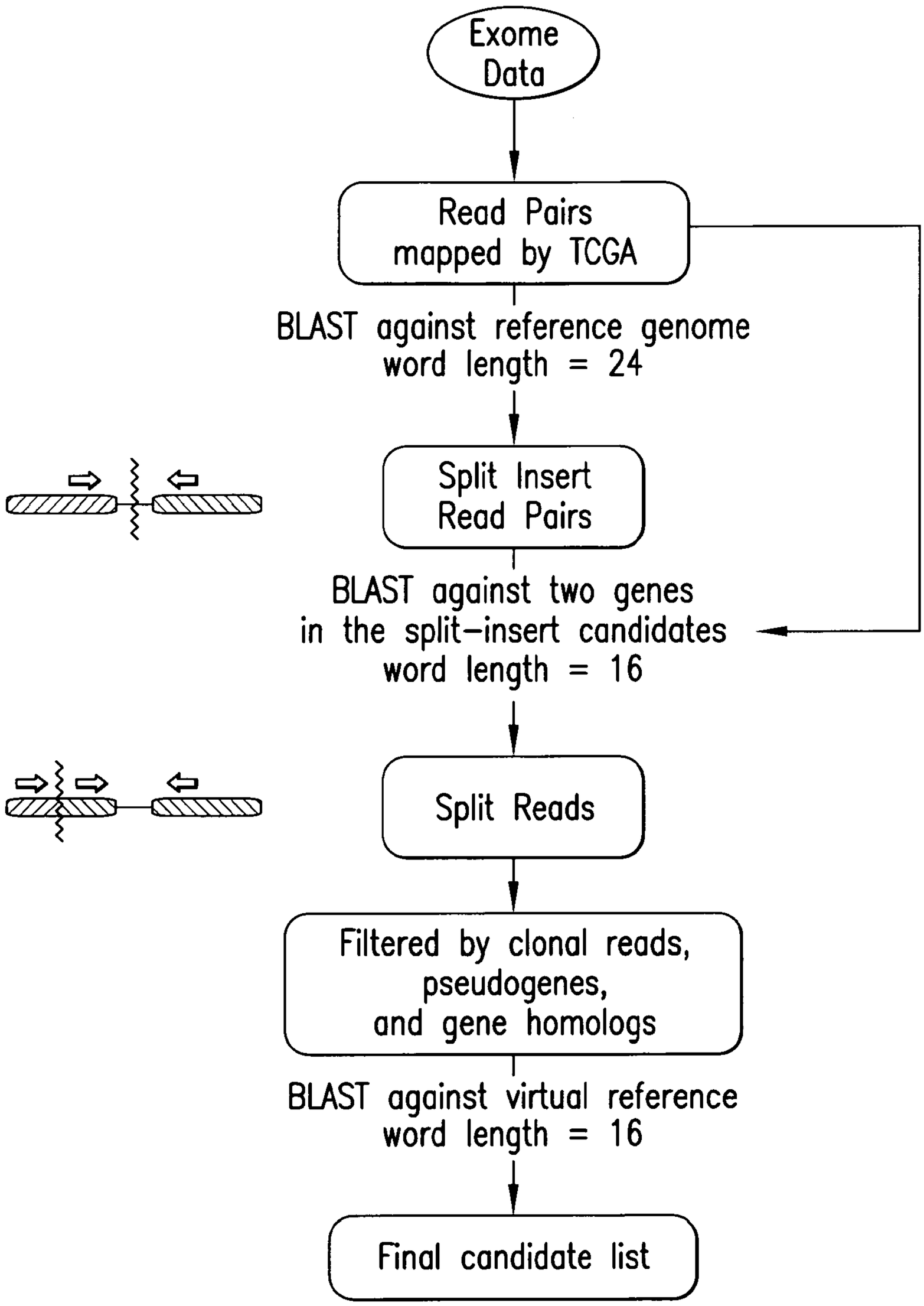

Exome Data
Read Pairs mapped by TCGA
BLAST against reference genome word length = 24
Split Insert Read Pairs
BLAST against two genes in the split-insert candidates word length = 16
Split Reads
Filtered by clonal reads, pseudogenes, and gene homologs
BLAST against virtual reference word length = 16
Final candidate list

FIG.8 - CONT.

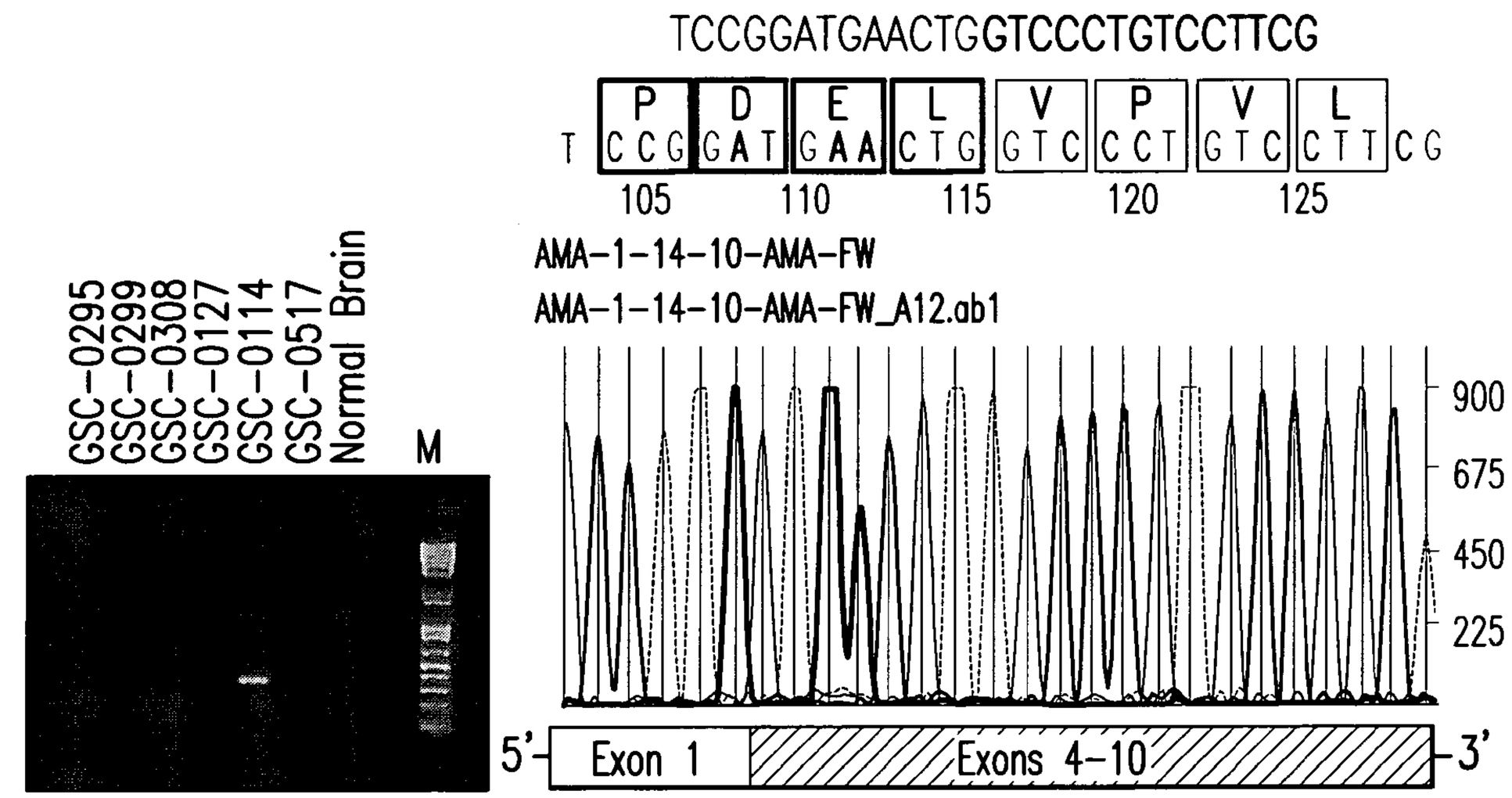

POLR2A-WRAP53 GSC-0114
TCCGGATGAACTGGTCCCTGTCCTTCG
T CCG GAT GAA CTG GTC CCT GTC CTT CG
P D E L V P V L
105 110 115 120 125
AMA-1-14-10-AMA-FW
AMA-1-14-10-AMA-FW_A12.ab1
GSC-0295
GSC-0299
GSC-0308
GSC-0127
GSC-0114
GSC-0517
Normal Brain
M
900
675
450
225
5'
Exon 1
Exons 4-10
3'

FIG.9A

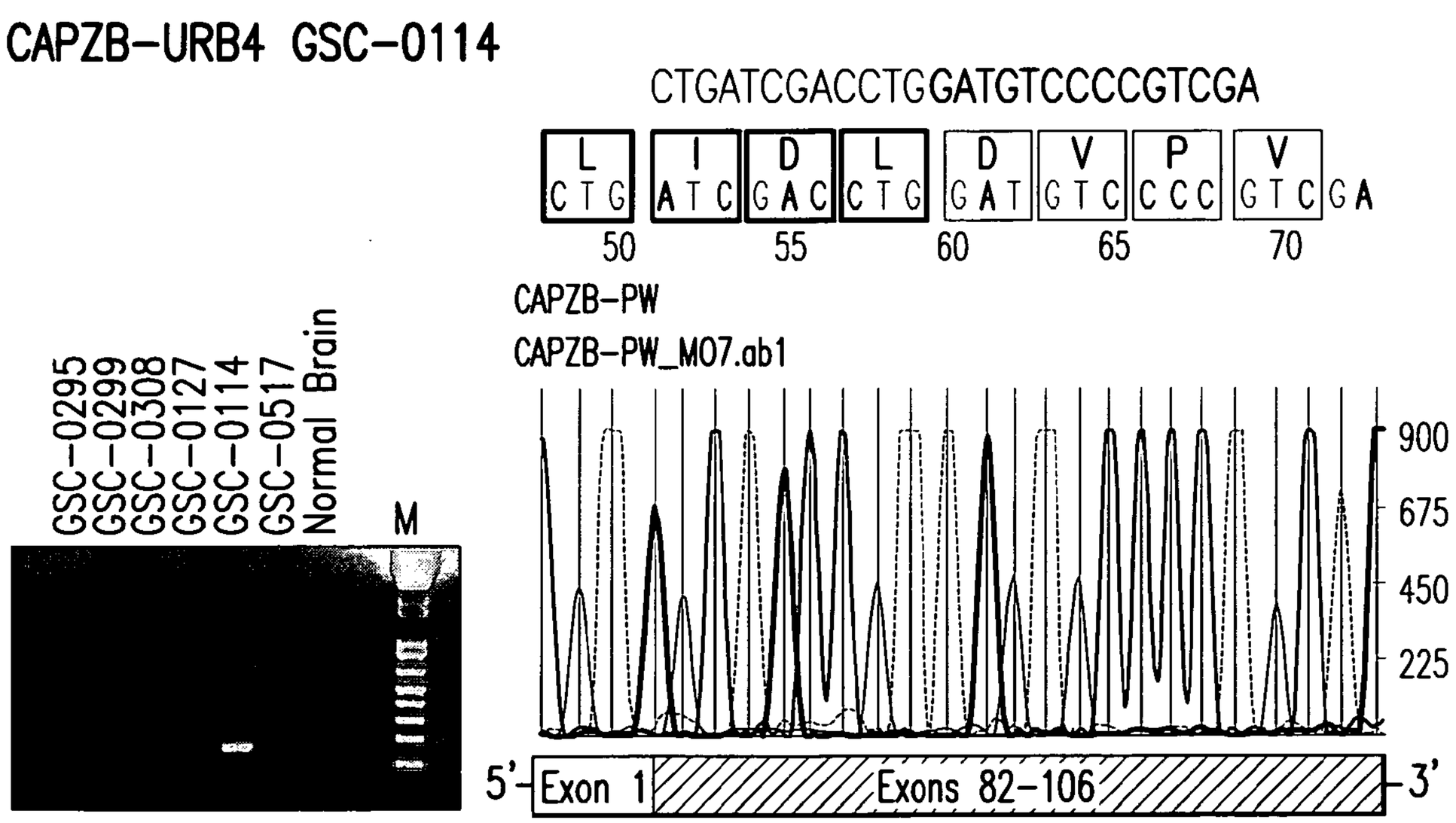

CAPZB-URB4 GSC-0114
CTGATCGACCTGGATGTCCCCGTCGA
L I D L D V P V
CTG ATC GAC CTG GAT GTC CCC GTC GA
50 55 60 65 70
CAPZB-PW
CAPZB-PW_M07.ab1
GSC-0295
GSC-0299
GSC-0308
GSC-0127
GSC-0114
GSC-0517
Normal Brain
M
900
675
450
225
5'
Exon 1
Exons 82-106
3'

FIG.9B

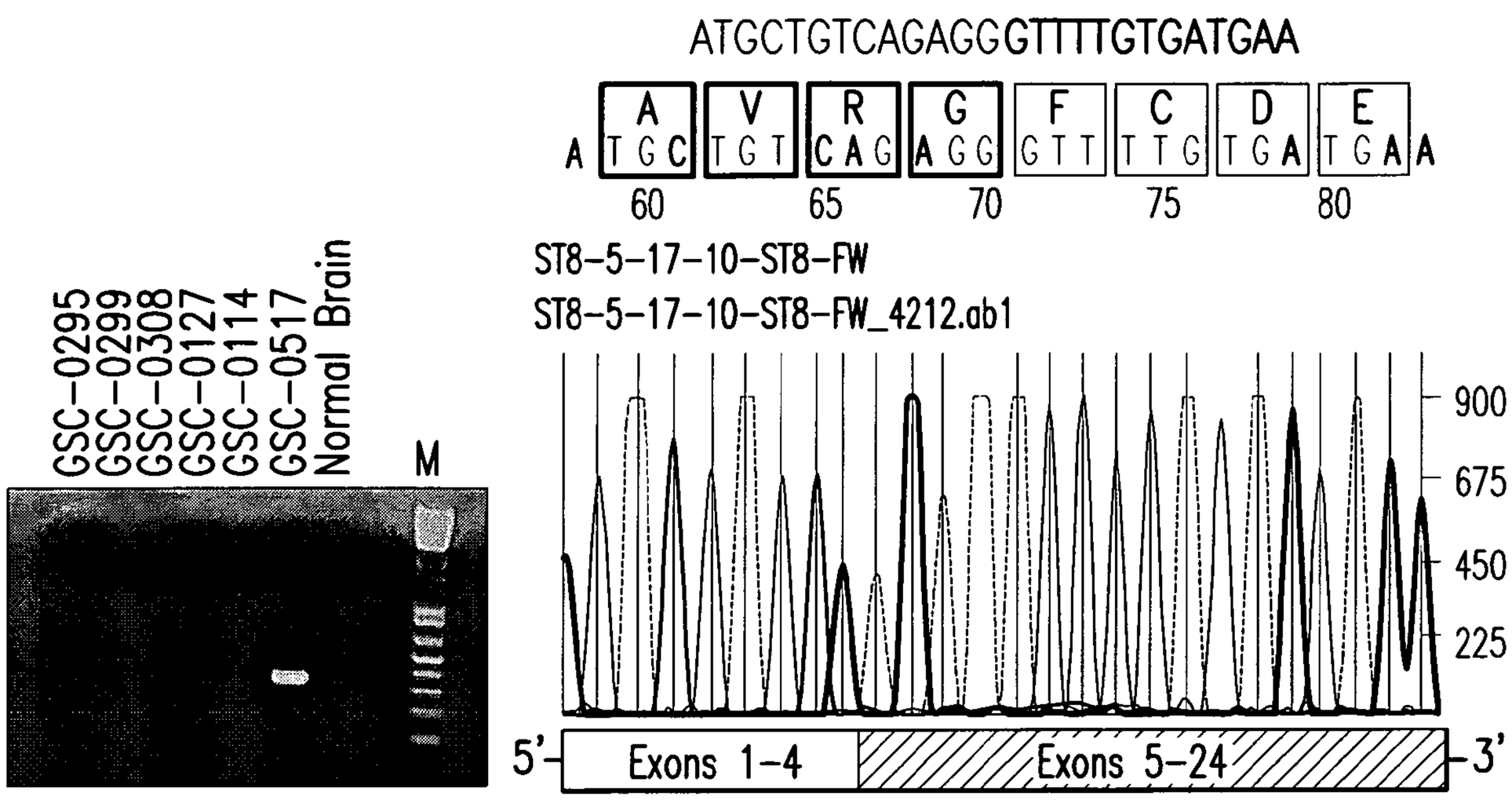

ST8SIA4-PAM GSC-0517
ATGCTGTCAGAGGGTTTTGTGATGAA
A
A T G C
V T G T
R C A G
G A G G
F G T T
C T T G
D T G A
E T G A A
60
65
70
75
80
ST8-5-17-10-ST8-FW
ST8-5-17-10-ST8-FW_4212.ab1
GSC-0295
GSC-0299
GSC-0308
GSC-0127
GSC-0114
GSC-0517
Normal Brain
M
900
675
450
225
5'
Exons 1-4
Exons 5-24
3'

FIG.9C

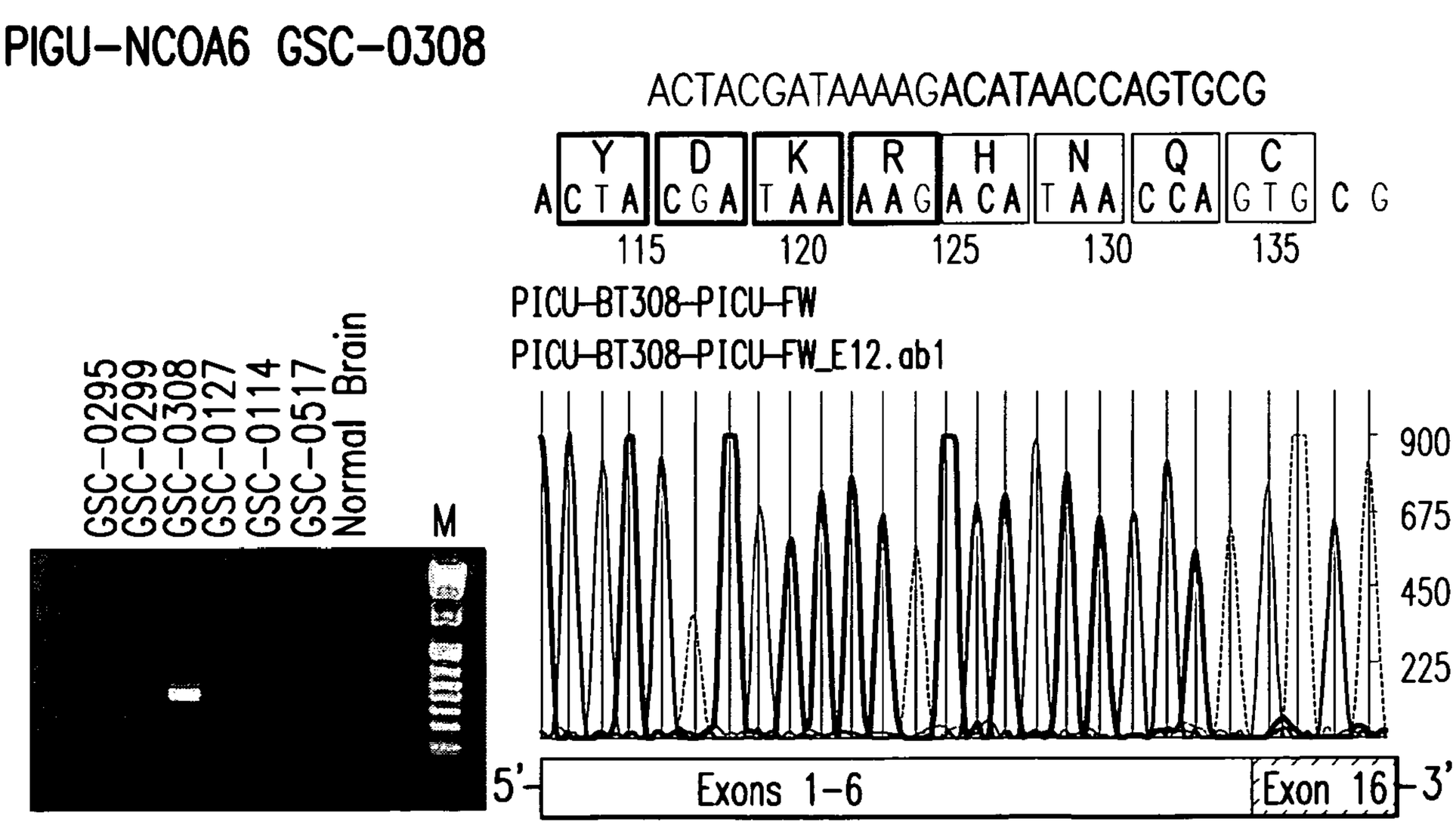

PIGU-NCOA6 GSC-0308
ACTACGATAAAAGACATAACCAGTGCG
A
Y C T A
D C G A
K T A A
R A A G
H A C A
N T A A
Q C C A
C G T G
C G
115
120
125
130
135
PICU-BT308-PICU-FW
PICU-BT308-PICU-FW_E12.ab1
GSC-0295
GSC-0299
GSC-0308
GSC-0127
GSC-0114
GSC-0517
Normal Brain
M
900
675
450
225
5'
Exons 1-6
Exon 16
3'

FIG.9D

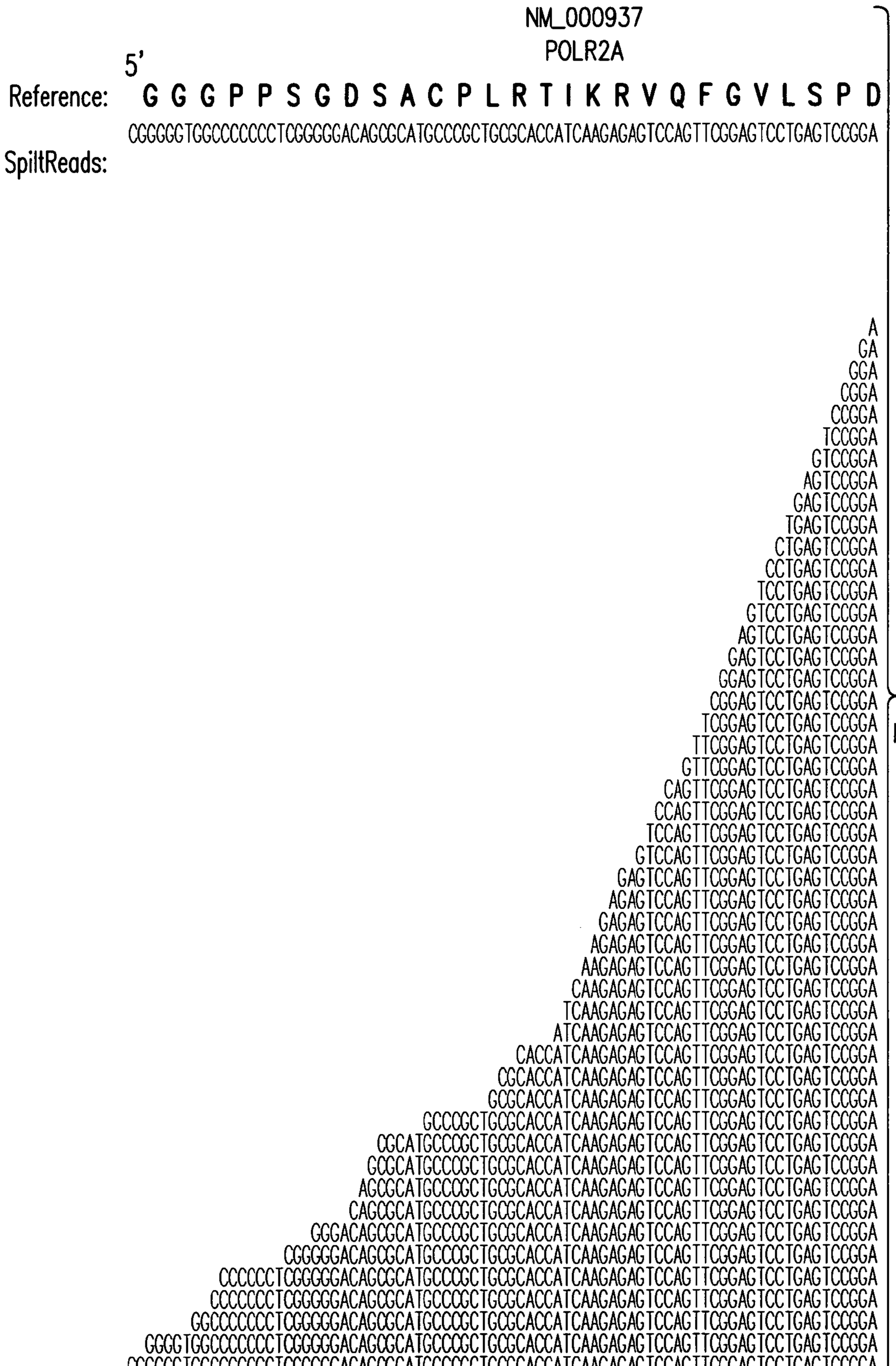

NM_000937
POLR2A
5'
Reference: G G G P P S G D S A C P L R T I K R V Q F G V L S P D
CGGGGGTGGCCCCCCCTCGGGGGACAGCGCATGCCCGCTGCGCACCATCAAGAGAGTCCAGTTCGGAGTCCTGAGTCCGGA
SpiltReads:
A
GA
GGA
CGGA
CCGGA
TCCGGA
GTCCGGA
AGTCCGGA
GAGTCCGGA
TGAGTCCGGA
CTGAGTCCGGA
CCTGAGTCCGGA
TCCTGAGTCCGGA
GTCCTGAGTCCGGA
AGTCCTGAGTCCGGA
GAGTCCTGAGTCCGGA
GGAGTCCTGAGTCCGGA
CGGAGTCCTGAGTCCGGA
TCGGAGTCCTGAGTCCGGA
TTCGGAGTCCTGAGTCCGGA
GTTCGGAGTCCTGAGTCCGGA
CAGTTCGGAGTCCTGAGTCCGGA
CCAGTTCGGAGTCCTGAGTCCGGA
TCCAGTTCGGAGTCCTGAGTCCGGA
GTCCAGTTCGGAGTCCTGAGTCCGGA
GAGTCCAGTTCGGAGTCCTGAGTCCGGA
AGAGTCCAGTTCGGAGTCCTGAGTCCGGA
GAGAGTCCAGTTCGGAGTCCTGAGTCCGGA
AGAGAGTCCAGTTCGGAGTCCTGAGTCCGGA
AAGAGAGTCCAGTTCGGAGTCCTGAGTCCGGA
CAAGAGAGTCCAGTTCGGAGTCCTGAGTCCGGA
TCAAGAGAGTCCAGTTCGGAGTCCTGAGTCCGGA
ATCAAGAGAGTCCAGTTCGGAGTCCTGAGTCCGGA
CACCATCAAGAGAGTCCAGTTCGGAGTCCTGAGTCCGGA
CGCACCATCAAGAGAGTCCAGTTCGGAGTCCTGAGTCCGGA
GCGCACCATCAAGAGAGTCCAGTTCGGAGTCCTGAGTCCGGA
GCCCGCTGCGCACCATCAAGAGAGTCCAGTTCGGAGTCCTGAGTCCGGA
CGCATGCCCGCTGCGCACCATCAAGAGAGTCCAGTTCGGAGTCCTGAGTCCGGA
GCGCATGCCCGCTGCGCACCATCAAGAGAGTCCAGTTCGGAGTCCTGAGTCCGGA
AGCGCATGCCCGCTGCGCACCATCAAGAGAGTCCAGTTCGGAGTCCTGAGTCCGGA
CAGCGCATGCCCGCTGCGCACCATCAAGAGAGTCCAGTTCGGAGTCCTGAGTCCGGA
GGGACAGCGCATGCCCGCTGCGCACCATCAAGAGAGTCCAGTTCGGAGTCCTGAGTCCGGA
CGGGGGACAGCGCATGCCCGCTGCGCACCATCAAGAGAGTCCAGTTCGGAGTCCTGAGTCCGGA
CCCCCCTCGGGGGACAGCGCATGCCCGCTGCGCACCATCAAGAGAGTCCAGTTCGGAGTCCTGAGTCCGGA
CCCCCCCTCGGGGGACAGCGCATGCCCGCTGCGCACCATCAAGAGAGTCCAGTTCGGAGTCCTGAGTCCGGA
GGCCCCCCCTCGGGGGACAGCGCATGCCCGCTGCGCACCATCAAGAGAGTCCAGTTCGGAGTCCTGAGTCCGGA
GGGGTGGCCCCCCCTCGGGGGACAGCGCATGCCCGCTGCGCACCATCAAGAGAGTCCAGTTCGGAGTCCTGAGTCCGGA
CGGGGGTGGCCCCCCCTCGGGGGACAGCGCATGCCCGCTGCGCACCATCAAGAGAGTCCAGTTCGGAGTCCTGAGTCCGGA
Cont.
On
FIG.9E-2

FIG.9E-1

NM_001143990
WRAP53

479 798 3'

E L V P V L R M V E G D T I Y D Y C W Y S L M S S A Q P D T S X

TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTCTGATGTCCTCAGCCCAGCCAGACACCTCCT

TG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTCTGATGTCCTCAGCCCAGCCAGACACCTCCT
CTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTCTGATGTCCTCAGCCCAGCCAGACACCTCC
ACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTCTGATGTCCTCAGCCCAGCCAGACACCAC
AACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTCTGATGTCCTCAGCCCAGCCAGACACCT
GAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTCTGATGTCCTCAGCCCAGCCAGACACC
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTCTGATGTCCTCAGCCCAGCCAGACAC
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTCTGATGTCCTCAGCCCAGCCAGACACC
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTCTGATGTCCTCAGCCCAGCCAGAC
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTCTGATGTCCTCAGCCCAGCCAGA
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTCTGATGTCCTCAGCCCAGCCAG
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTCTGATGTCCTCAGCCCAGCCA
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTCTGATGTCCTCAGCCCAGCC
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTCTGATGTCCTCAGCCCAGC
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTCTGATGTCCTCAGCCCAG
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTCTGATGTCCTCAGCCCA
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTCTGATGTCCTCAGCCC
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTCTGATGTCCTCAGCC
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTCTGATGTCCTCAGC
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTCTGATGTCCTCAG
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTCTGATGTCCTCA
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTCTGATGTCCTC
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTCTGATGTCCT
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTCTGATGTCC
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTCTGATGTC
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTCTGATGT
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTCTGATG
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTCTGAT
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTCTG
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTCT
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCTC
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATTCT
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTATT
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTAT
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGTA
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTGGT
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCCGG
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCTG
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGCT
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTACTGC
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGATTA
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGAT
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCATCTATGA
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGATACCA
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTGA
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGTG
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGGT
TGAACTG GTCCCTGTCCTTCGAATGGTGGAAGG
TGAACTG GTCCCTGTCCTTCGAATGGTGG
TGAACTG GTCCCTGTCCTTCGAATGG
TGAACTG GTCCCTGTCCTT
TGAACTG GTCCCTGTCCT
TGAACTG GTCCCTGTC
TGAACTG GTCC
TGAACTG GT

Cont. From FIG.9E-1

FIG.9E-2

NM_001206540
CAPZB

5'

Reference: Q L D C A L D L M R R L P P Q Q I E K N L S D L I D

GCAGC-TGGACTGTGCCTTGGACCTAATGAGGCGCCTGCCTCCCCAGCAAATCGAGAAAAACCTCAGCGACCTGATCGAC

SpiltReads:

C
AC
GCC
CCAC
GCGAC
GATCGAC
ACCTGATCGAC
CGACCTGATCGAC
CTCAGCGACCTGATCGAC
AAAAACCTCAGCGACCTGATCGAC
GAAAAACCTCAGCGACCTGATCGAC
AGAAAAACCTCAGCGACCTGATCGAC
CGAGAAAAACCTCAGCGACCTGATCGAC
TCGAGAAAAACCTCAGCGACCTGATCGAC
ATCGAGAAAAACCTCAGCGACCTGATCGAC
AATCGAGAAAAACCTCAGCGACCTGATCGAC
AAATCGAGAAAAACCTCAGCGACCTGATCGAC
CAAATCGAGAAAAACCTCAGCGACCTGATCGAC
GCAAATCGAGAAAAACCTCAGCGACCTGATCGAC
AGCAAATCGAGAAAAACCTCAGCGACCTGATCGAC
CAGCAAATCGAGAAAAACCTCAGCGACCTGATCGAC
CCCCAGCAAATCGAGAAAAACCTCAGCGACCTGATCGAC
CTCCCCAGCAAATCGAGAAAAACCTCAGCGACCTGATCGAC
CCTCCCCAGCAAATCGAGAAAAACCTCAGCGACCTGATCGAC
GCCTCCCCAGCAAATCGAGAAAAACCTCAGCGACCTGATCGAC
TGCCTCCCCAGCAAATCGAGAAAAACCTCAGCGACCTGATCGAC
CTGCCTCCCCAGCAAATCGAGAAAAACCTCAGCGACCTGATCGAC
CCTGCCTCCCCAGCAAATCGAGAAAAACCTCAGCGACCTGATCGAC
CGCCTGCCTCCCCAGCAAATCGAGAAAAACCTCAGCGACCTGATCGAC
GGCGCCTGCCTCCCCAGCAAATCGAGAAAAACCTCAGCGACCTGATCGAC
GCGGCGCCTGCCTCCCCAGCAAATCGAGAAAAACCTCAGCGACCTGATCGAC
TGAGGCGCCTGCCTCCCCAGCAAATCGAGAAAAACCTCAGCGACCTGATCGAC
ATGAGGCGCCTGCCTCCCCAGCAAATCGAGAAAAACCTCAGCGACCTGATCGAC
AATGAGGCGCCTGCCTCCCCAGCAAATCGAGAAAAACCTCAGCGACCTGATCGAC
TAATGAGGCGCCTGCCTCCCCAGCAAATCGAGAAAAACCTCAGCGACCTGATCGAC
GACCTAATGAGGCGCCTGCCTCCCCAGCAAATCGAGAAAAACCTCAGCGACCTGATCGAC
TTGGACCTAATGAGGCGCCTGCCTCCCCAGCAAATCGAGAAAAACCTCAGCGACCTGATCGAC
CTTGGACCTAATGAGGCGCCTGCCTCCCCAGCAAATCGAGAAAAACCTCAGCGACCTGATCGAC
GCCTTGGACCTAATGAGGCGCCTGCCTCCCCAGCAAATCGAGAAAAACCTCAGCGACCTGATCGAC
GTGCCTTGGACCTAATGAGGCGCCTGCCTCCCCAGCAAATCGAGAAAAACCTCAGCGACCTGATCGAC
ACTGTGCCTTGGACCTAATGAGGCGCCTGCCTCCCCAGCAAATCGAGAAAAACCTCAGCGACCTGATCGAC
GACTGTGCCTTGGACCTAATGAGGCGCCTGCCTCCCCAGCAAATCGAGAAAAACCTCAGCGACCTGATCGAC
CATGGACTGTGCCTTGGACCTAATGAGGCGCCTGCCTCCCCAGCAAATCGAGAAAAACCTCAGCGACCTGATCGAC
GC-TGGACTGTGCCTTGGACCTAATGAGGCGCCTGCCTCCCCAGCAAATCGAGAAAAACCTCAGCGACCTGATCGAC
AGC-TGGACTGTGCCTTGGACCTAATGAGGCGCCTGCCTCCCCAGCAAATCGAGAAAAACCTCAGCGACCTGATCGAC
GCAGC-TGGACTGTGCCTTGGACCTAATGAGGCGCCTGCCTCCCCAGCAAATCGAGAAAAACCTCAGCGACCTGATCGAC

Cont. On FIG.9E-4

FIG.9E-3

NM_020765
UBR4

227 12110 3'

L D V P V E A L T T V K P Y C N E I H A Q A Q L W L K R D P

CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTGCAATGAGATCCATGCCCAGGCTCAACTGTGGCTCAAGAGAGACCCCA

TG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTGCAATGAGATCCATGCCCAGGCTCAACTGTGGCTCAAGAGAGACCCCA
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTGCAATGAGATCCATGCCCAGGCTCAACTGTGGCTCAAGAGAGACCCC
AAG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTGCAATGAGATCCATGCCCAGGCTCAACTGTGGCTCAAGAGAGACCT
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTGCAATGAGATCCATGCCCAGGCTCAACTGTGGCTCAAGAGAGACC
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTGCAATGAGATCCATGCCCAGGCTCAACTGTGGCTCAAGAGAGAC
CTT GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTGCAATGAGATCCATGCCCAGGCTCAACTGTGGCTCAAGAGAGA
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTGCAATGAGATCCATGCCCAGGCTCAACTGTGGCTCAAGAGAG
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTGCAATGAGATCCATGCCCAGGCTCAACTGTGGCTCAAGAG
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTGCAATGAGATCCATGCCCAGGCTCAACTGTGGCTCA
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTGCAATGAGATCCATGCCCAGGCTCAACTGTGGCT
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTGCAATGAGATCCATGCCCAGGCTCAACTG
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTGCAATGAGATCCATGCCCAGGCTC
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTGCAATGAGATCCATGCCCAGGC
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTGCAATGAGATCCATGCCCAGG
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTGCAATGAGATCCATGCCCA
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTGCAATGAGATCCATGCCC
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTGCAATGAGATCCATGCC
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTGCAATGAGATCCATGC
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTGCAATGAGATCCATG
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTGCAATGAGATCCAT
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTGCAATGAGATCCA
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTGCAATGAGATCC
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTGCAATGAGATC
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTGCAATGAG
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTGCAATG
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTGCAAT
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTGCAA
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTGCA
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTGC
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATACTG
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCATAC
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCCAT
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGCC
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAGC
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAAG
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGAA
CTG GATGTCCCCGTTGAGGCCCTCACCACGTGA
CTG GATGTCCCCGTTGAGGCCCTCACCACG
CTG GATGTCCCCGTTGAGGCCCTCACC
CTG GATGTCCCCGTTGAGGCCCTCAC
CTG GATGTCCCCGTTGAGGCCCTC
CTG GATGTCCCCGTTGAGGCCC
CTG GATGTCCCCGTTGAGG
CTG GATGTCCCCGTCGAG
CTG GATGTCCCCGA
CTG GATGTCCCCCG
CTG GATGTCCCCGG
CTG GATGTCCCC

Cont. From FIG.9E-3

FIG.9E-4

NM_005668
ST8SIA4

5'

Reference: **E W V N A L I L K N K L K V R T A Y P S L R L I H A**

GGAGTGGGTTAATGCATTAATCCTTAAGAATAAACTGAAAGTGCGAACTGCCTATCCGTCATTGAGACTTATTCATGC

SpiltReads:

GC
TGC
ATGC
CATGC
TCATGC
TATTCATGC
CTTATTCATGC
ACTTATTCATGC
GAGACTTATTCATGC
TGAGACTTATTCATGC
TTGAGACTTATTCATGC
GTCATTGAGACTTATTCATGC
CCGTCATTGAGACTTATTCATGC
TATCCGTCATTGAGACTTATTCATGC
CTATCCGTCATTGAGACTTATTCATGC
CCTATCCGTCATTGAGACTTATTCATGC
GCCTATCCGTCATTGAGACTTATTCATGC
TGAAAGTGCGAACTGCCTATCCGTCATTGAGACTTATTCATGC
CTGAAAGTGCGAACTGCCTATCCGTCATTGAGACTTATTCATGC
AAACTGAAAGTGCGAACTGCCTATCCGTCATTGAGACTTATTCATGC
TAAACTGAAAGTGCGAACTGCCTATCCGTCATTGAGACTTATTCATGC
ATTAATCCTTAAGAATAAACTGAAAGTGCGAACTGCCTATCCGTCATTGAGACTTATTCATGC
GCATTAATCCTTAAGAATAAACTGAAAGTGCGAACTGCCTATCCGTCATTGAGACTTATTCATGC
GGAGTGGGTTAATGCATTAATCCTTAAGAATAAACTGAAAGTGCGAACTGCCTATCCGTCATTGAGACTTATTCATGC

Cont. On FIG.9E-6

FIG.9E-5

NM_000919

1125 730 PAM 3'

V R G F C D E G T C T D K A N I L Y A W A R N A P P T R L P K G X

Cont. From FIG.9E-5

TGTCAGAGG GTTTTGTGATGAAGGAACCTGTACAGATAAAGCCAATATTCTGTATGCCTGGGCGAGAAATGCTCCCCCTACCGGCTCCCCAAAGGTG

G GTTTTGTGATGAAGGAACCTGTACAGATAAAGCCAATATTCTGTATGCCTGGGCGAGAAATGCTCCCCCTACCGGCTCCCCAAAGGTG

TG GTTTTGTGATGAAGGAACCTGTACAGATAAAGCCAATATTCTGTATGCCTGGGCGAGAAATGCTCCCCCTACCGGCTCCCCAAAGGT

CTG GTTTTGTGATGAAGGAACCTGTACAGATAAAGCCAATATTCTGTATGCCTGGGCGAGAAATGCTCCCCCTACCGGCTCCCCAAAGG

GAGG GTTTTGTGATGAAGGAACCTGTACAGATAAAGCCAATATTCTGTATGCCTGGGCGAGAAATGCTCCCCCTACCGGCTCCCCAAAG

GTTA-CTG GTTTTGTGATGAAGGAACCTGTACAGATAAAGCCAATATTCTGTATGCCTGGGCGAGAAATGCTCCCCCTACCGGCTCCCCAA

TGTCAGAGG GTTTTGTGATGAAGGAACCTGTACAGATAAAGCCAATATTCTGTATGCCTGGGCGAGAAATGCTCCCCCTACCGGCTC

TGTCAGAGG GTTTTGTGATGAAGGAACCTGTACAGATAAAGCCAATATTCTGTATGCCTGGGCGAGAAATGCTCCCCCTACCGGC-CC

TGTCAGAGG GTTTTGTGATGAAGGAACCTGTACAGATAAAGCCAATATTCTGTATGCCTGGGCGAGAAATGCTCCCCCTACCGGC

TGTCAGAGG GTTTTGTGATGAAGGAACCTGTACAGATAAAGCCAATATTCTGTATGCCTGGGCGAGAAATGCTCCCCCTACCGG

TGTCAGAGG GTTTTGTGATGAAGGAACCTGTACAGATAAAGCCAATATTCTGTATGCCTGGGCGAGAAATGCTCCCCCTACCG

TGTCAGAGG GTTTTGTGATGAAGGAACCTGTACAGATAAAGCCAATATTCTGTATGCCTGGGCGAGAAATGCTCCCCCTAC

TGTCAGAGG GTTTTGTGATGAAGGAACCTGTACAGATAAAGCCAATATTCTGTATGCCTGGGCGAGAAATGCTCCCCCT

TGTCAGAGG GTTTTGTGATGAAGGAACCTGTACAGATAAAGCCAATATTCTGTATGCCTGGGCGAGAAATGCTCCCCC

TGTCAGAGG GTTTTGTGATGAAGGAACCTGTACAGATAAAGCCAATATTCTGTATGCCTGGGCGAGAAATGCTCC

TGTCAGAGG GTTTTGTGATGAAGGAACCTGTACAGATAAAGCCAATATTCTGTATGCCTGGGCGAGAAATGCTC

TGTCAGAGG GTTTTGTGATGAAGGAACCTGTACAGATAAAGCCAATATTCTGTATGCCTGGGCGAGAAATGCT

TGTCAGAGG GTTTTGTGATGAAGGAACCTGTACAGATAAAGCCAATATTCTGTATGCCTGGGCGAGAAA

TGTCAGAGG GTTTTGTGATGAAGGAACCTGTACAGATAAAGCCAATATTCTGTATGCCTGGGCGAGA

TGTCAGAGG GTTTTGTGATGAAGGAACCTGTACAGATAAAGCCAATATTCTGTATGCCTGGGCG

TGTCAGAGG GTTTTGTGATGAAGGAACCTGTACAGATAAAGCCAATATTCTGTATGCCTGGGC

TGTCAGAGG GTTTTGTGATGAAGGAACCTGTACAGATAAAGCCAATATTCTGTATGCCTGGG

TGTCAGAGG GTTTTGTGATGAAGGAACCTGTACAGATAAAGCCAATATTCTGTATGCCTGG

TGTCAGAGG GTTTTGTGATGAAGGAACCTGTACAGATAAAGCCAATA

TGTCAGAGG GTTTTGTGATGAAGGAACCTGTACAGATAAAGCCAAT

TGTCAGAGG GTTTTGTGATGAAGGAACCTGTACAGATAAAGCC

TGTCAGAGG GTTTTGTGATGAAGGAACCTGTACAGATAAAGC

TGTCAGAGG GTTTTGTGATGAAGGAAC

TGTCAGAGG GTTTTGTGATGAAGGA

TGTCAGAGG GTT

FIG.9E-6

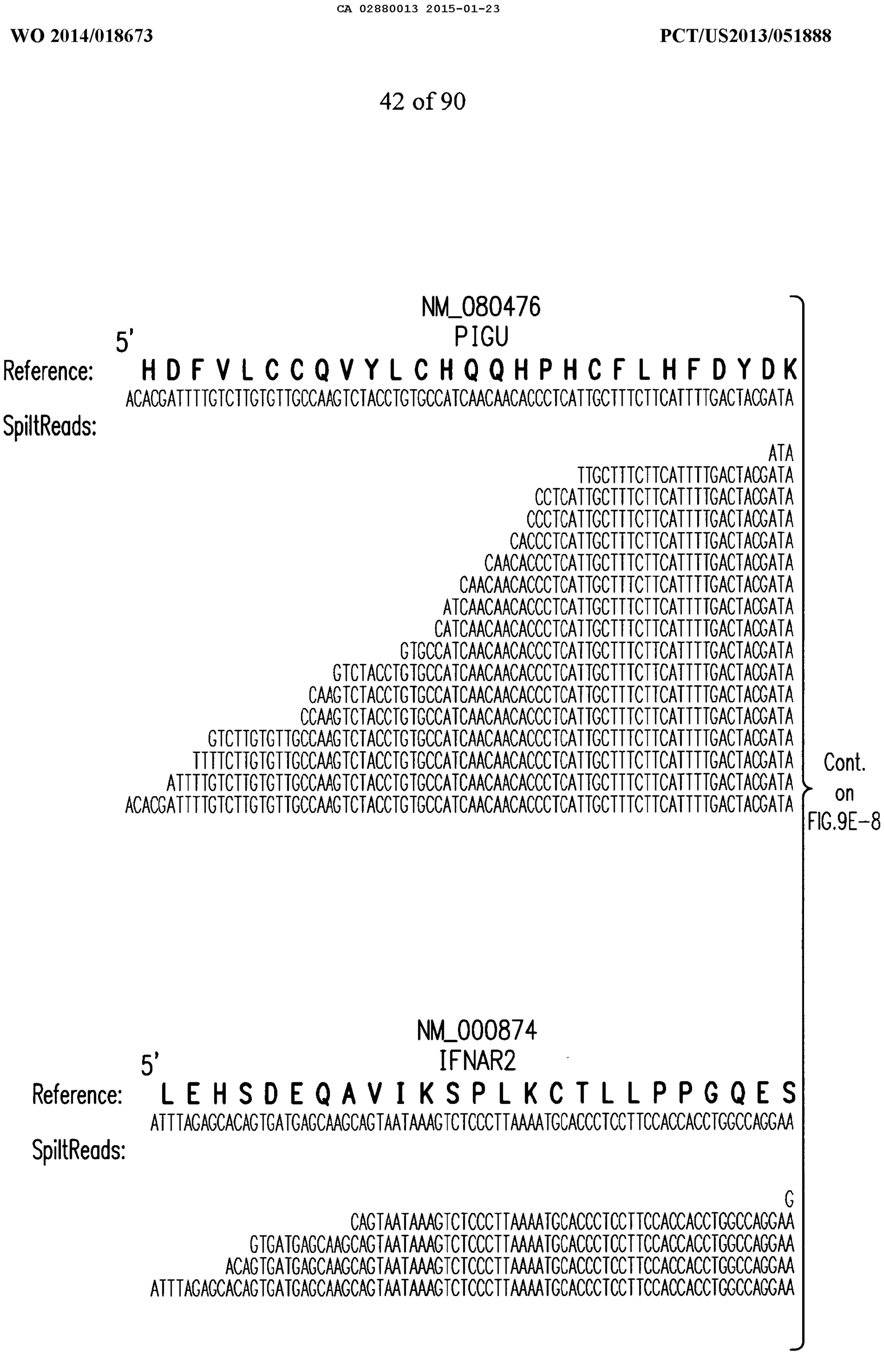

NM_080476
PIGU
5'
Reference:
H D F V L C C Q V Y L C H Q Q H P H C F L H F D Y D K
ACACGATTTTGTCTTGTGTTGCCAAGTCTACCTGTGCCATCAACAACACCCTCATTGCTTTCTTCATTTTGACTACGATA
SpiltReads:
ATA
TTGCTTTCTTCATTTTGACTACGATA
CCTCATTGCTTTCTTCATTTTGACTACGATA
CCCTCATTGCTTTCTTCATTTTGACTACGATA
CACCCTCATTGCTTTCTTCATTTTGACTACGATA
CAACACCCTCATTGCTTTCTTCATTTTGACTACGATA
CAACAACACCCTCATTGCTTTCTTCATTTTGACTACGATA
ATCAACAACACCCTCATTGCTTTCTTCATTTTGACTACGATA
CATCAACAACACCCTCATTGCTTTCTTCATTTTGACTACGATA
GTGCCATCAACAACACCCTCATTGCTTTCTTCATTTTGACTACGATA
GTCTACCTGTGCCATCAACAACACCCTCATTGCTTTCTTCATTTTGACTACGATA
CAAGTCTACCTGTGCCATCAACAACACCCTCATTGCTTTCTTCATTTTGACTACGATA
CCAAGTCTACCTGTGCCATCAACAACACCCTCATTGCTTTCTTCATTTTGACTACGATA
GTCTTGTGTTGCCAAGTCTACCTGTGCCATCAACAACACCCTCATTGCTTTCTTCATTTTGACTACGATA
TTTTCTTGTGTTGCCAAGTCTACCTGTGCCATCAACAACACCCTCATTGCTTTCTTCATTTTGACTACGATA
ATTTTGTCTTGTGTTGCCAAGTCTACCTGTGCCATCAACAACACCCTCATTGCTTTCTTCATTTTGACTACGATA
ACACGATTTTGTCTTGTGTTGCCAAGTCTACCTGTGCCATCAACAACACCCTCATTGCTTTCTTCATTTTGACTACGATA
Cont. on FIG.9E-8
NM_000874
IFNAR2
5'
Reference:
L E H S D E Q A V I K S P L K C T L L P P G Q E S
ATTTAGAGCACAGTGATGAGCAAGCAGTAATAAAGTCTCCCTTAAAATGCACCCTCCTTCCACCACCTGGCCAGGAA
SpiltReads:
G
CAGTAATAAAGTCTCCCTTAAAATGCACCCTCCTTCCACCACCTGGCCAGGAA
GTGATGAGCAAGCAGTAATAAAGTCTCCCTTAAAATGCACCCTCCTTCCACCACCTGGCCAGGAA
ACAGTGATGAGCAAGCAGTAATAAAGTCTCCCTTAAAATGCACCCTCCTTCCACCACCTGGCCAGGAA
ATTTAGAGCACAGTGATGAGCAAGCAGTAATAAAGTCTCCCTTAAAATGCACCCTCCTTCCACCACCTGGCCAGGAA

FIG.9E-7

NM_014071
NCOA6

729 6471 3'

R H N Q C G A I Q A K K I Q V N K Q D C D L I L G N V C X

AAAG ACATAACCAGTGCGGTGCAATCCAAGCGAAGAAAATCCAAGTAAACAAGCAGGACTGCGACTTGATACTTGGAAATGTGTGTG

AAAG ACATAACCAGTGCGGTGCAATCCAAGCGAAGAAAATCCAAGTAAACAAGCAGGACTGCGACTTGATACTTGGAAATGTGTGTG
AAAG ACATAACCAGTGCGGTGCAATCCAAGCGAAGAAAATCCAAGTAAACAAGCAGGACTGCGA
AAAG ACATAACCAGTGCGGTGCAATCCAAGCGAAGAAAATCCAAGTAAACAAGCAGGAC
AAAG ACATAACCAGTGCGGTGCAATCCAAGCGAAGAAAATCCAAGTAAACAAGCAGGA
AAAG ACATAACCAGTGCGGTGCAATCCAAGCGAAGAAAATCCAAGTAAACAAGCAG
AAAG ACATAACCAGTGCGGTGCAATCCAAGCGAAGAAAATCCAAGTAAACAAG
AAAG ACATAACCAGTGCGGTGCAATCCAAGCGAAGAAAATCCAAGTAAAC
AAAG ACATAACCAGTGCGGTGCAATCCAAGCGAAGAAAATCCAAGTAA
AAAG ACATAACCAGTGCGGTGCAATCCAAGCGAAGAAAATCCAAGTA
AAAG ACATAACCAGTGCGGTGCAATCCAAGCGAAGAAAATCCA
AAAG ACATAACCAGTGCGGTGCAATCCAAGCGAAG
AAAG ACATAACCAGTGCGGTGCAATCCAAGCG
AAAG ACATAACCAGTGCGGTGCAATCCAAGC
AAAG ACATAACCAGTGCGGT
AAAG ACATAACCAGTGCG
AAAG ACATAACCAGTGCG
AAAG ACATAACCAGT
AAAG ACATAA

Cont. From FIG.9E-7

NM_000628
IL10RB

1083 149 3'

A L G M V P P P E N V R M N S V N F K N I L Q W E S P A F X

TCAG CATTGGGAATGGTACCACCTCCCGAAAATGTCAGAATGAATTCTGTTAATTTCAAGAACATTCTACAGTGGGAGTCACCTGCTTTTGC

AG CATTGGGAATGGTACCACCTCCCGAAAATGTCAGAATGAATTCTGTTAATTTCAAGAACATTCTACAGTGGGAGTCACCTGCTTTTGC
TCAG CATTGGGAATGGTACCACCTCCCGAAAATGTCAGAATGAATTCTGTTAATTTCAAGAACATTCTACAGTGGGAGTCACCTGCTTT
TCAG CATTGGGAATGGTACCACCTCCCGAAAATGTCAGAA
TCAG CATTGGGAATGGTACCACCTCCCG
TCAG CATTGGGAATGGTACCACCAA
TCAG CATTGGGAATGG

FIG.9E-8

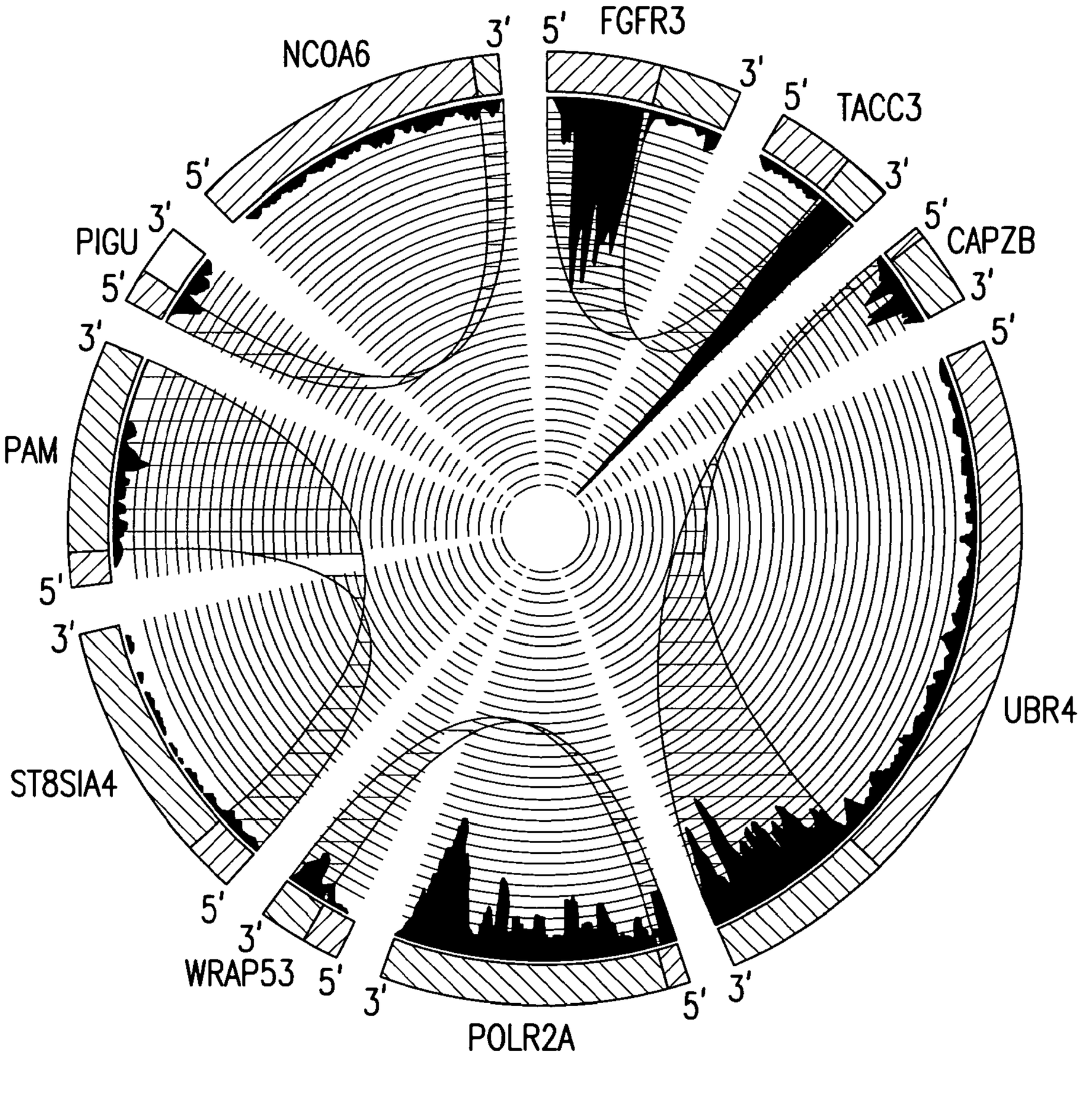
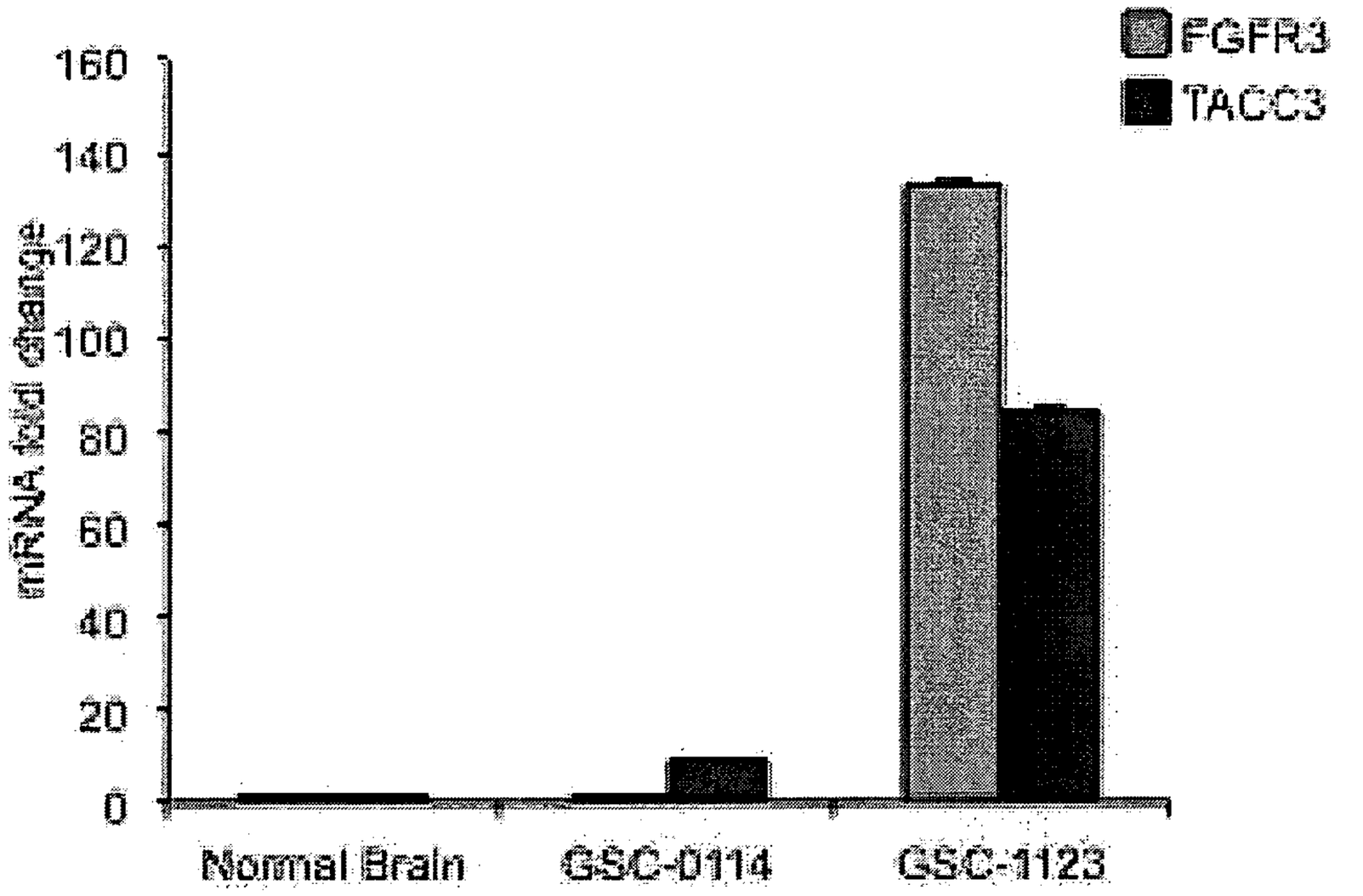

NCOA6
3'
5'
FGFR3
5'
3'
TACC3
5'
3'
CAPZB
5'
3'
PIGU
3'
5'
PAM
5'
3'
UBR4
5'
3'
ST8SIA4
5'
3'
WRAP53
5'
3'
POLR2A
5'
3'

FIG.10A


FGFR3
TACC3
mRNA fold change
160
140
120
100
80
60
40
20
0
Normal Brain
GSC-0114
GSC-1123

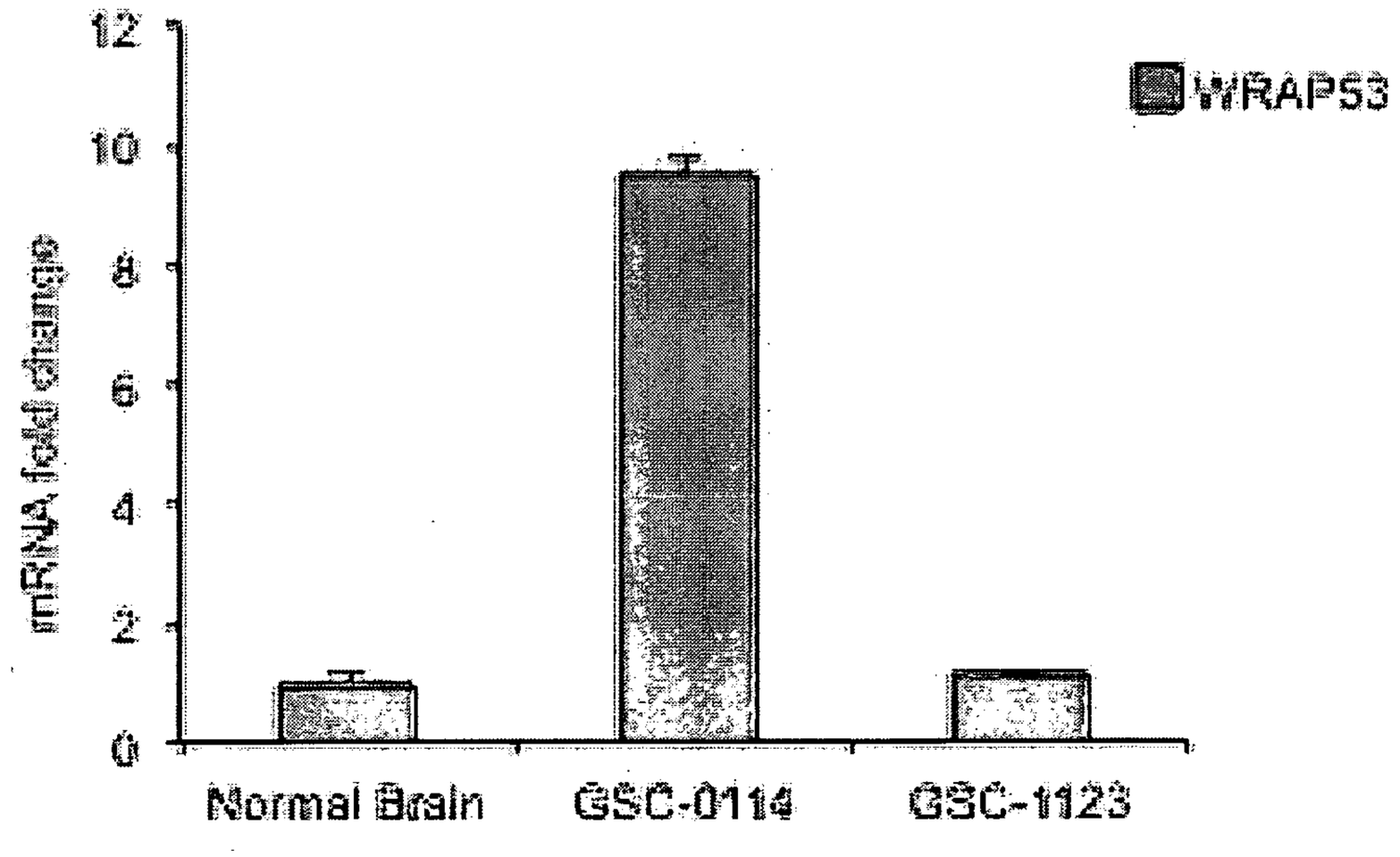

WRAP53
mRNA fold change
12
10
8
6
4
2
0
Normal Brain
GSC-0114
GSC-1123

FIG. 10B

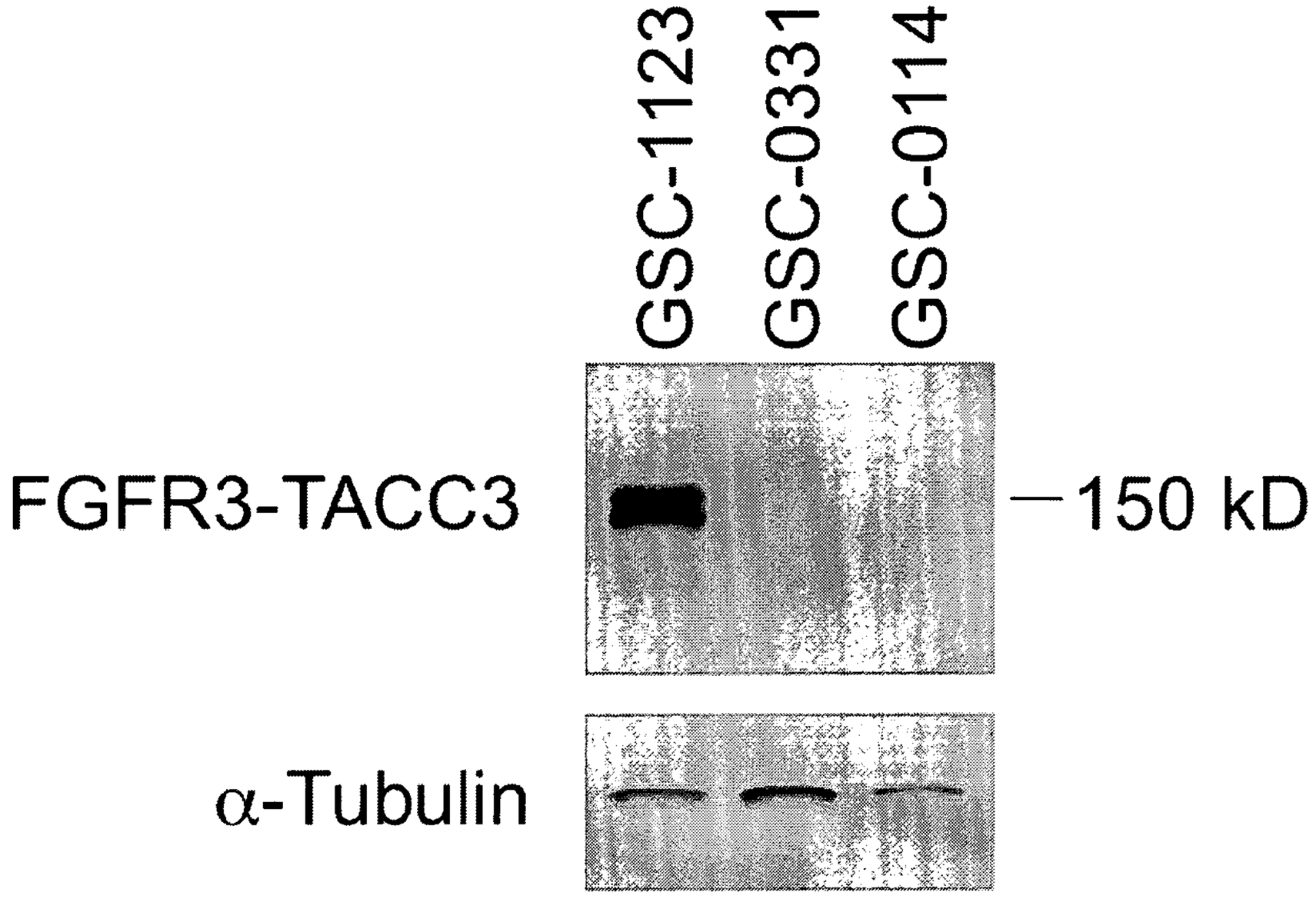


FIG. 10C

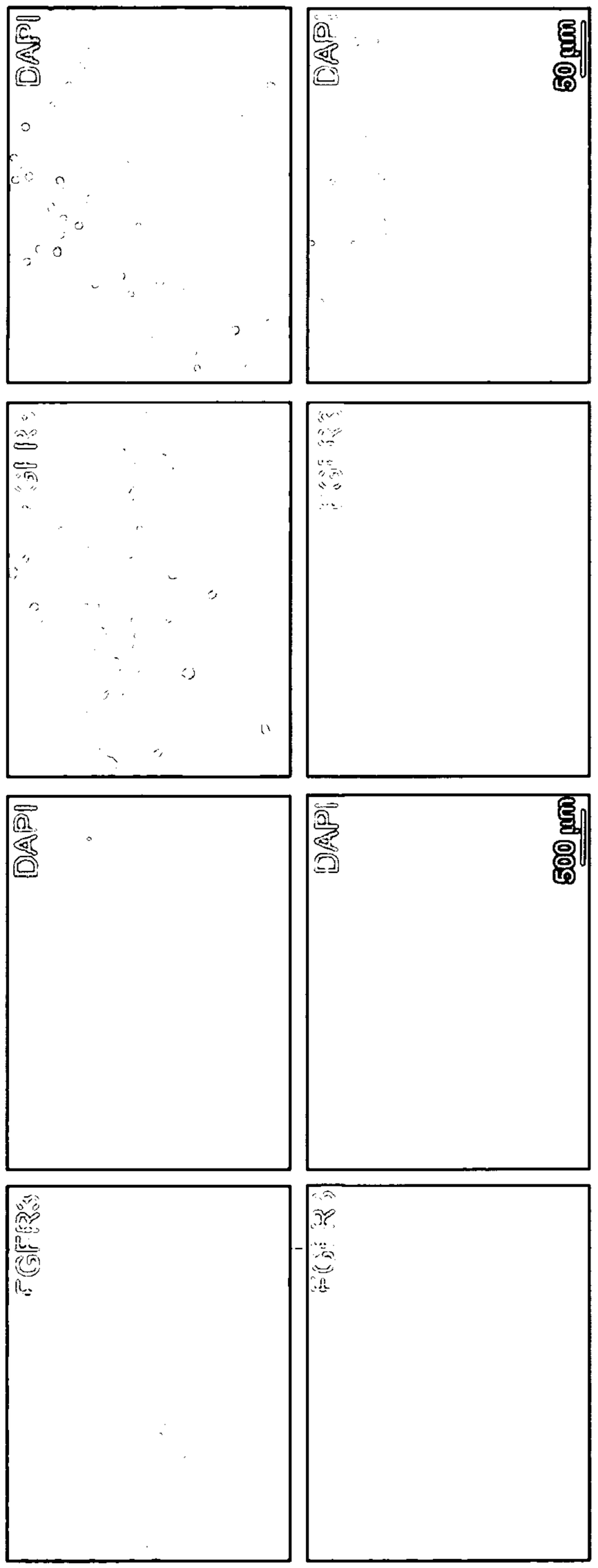


FIG. 10D

Seq	#	b	y	+1
A	1	72.0	–	13
Q	2	**200.1**	1270.7	12
A	3	271.1	1142.6	11
E	4	400.2	1071.6	10
A	5	**471.2**	942.6	9
L	6	**584.3**	871.5	8
A	7	**655.3**	758.4	7
L	8	**768.4**	687.4	6
Q	9	**896.5**	574.3	5
A	10	**967.5**	446.3	4
S	11	**1054.5**	375.2	3
L	12	**1167.6**	288.2	2
R	13	–	175.1	1

TACC3 Peptide 1

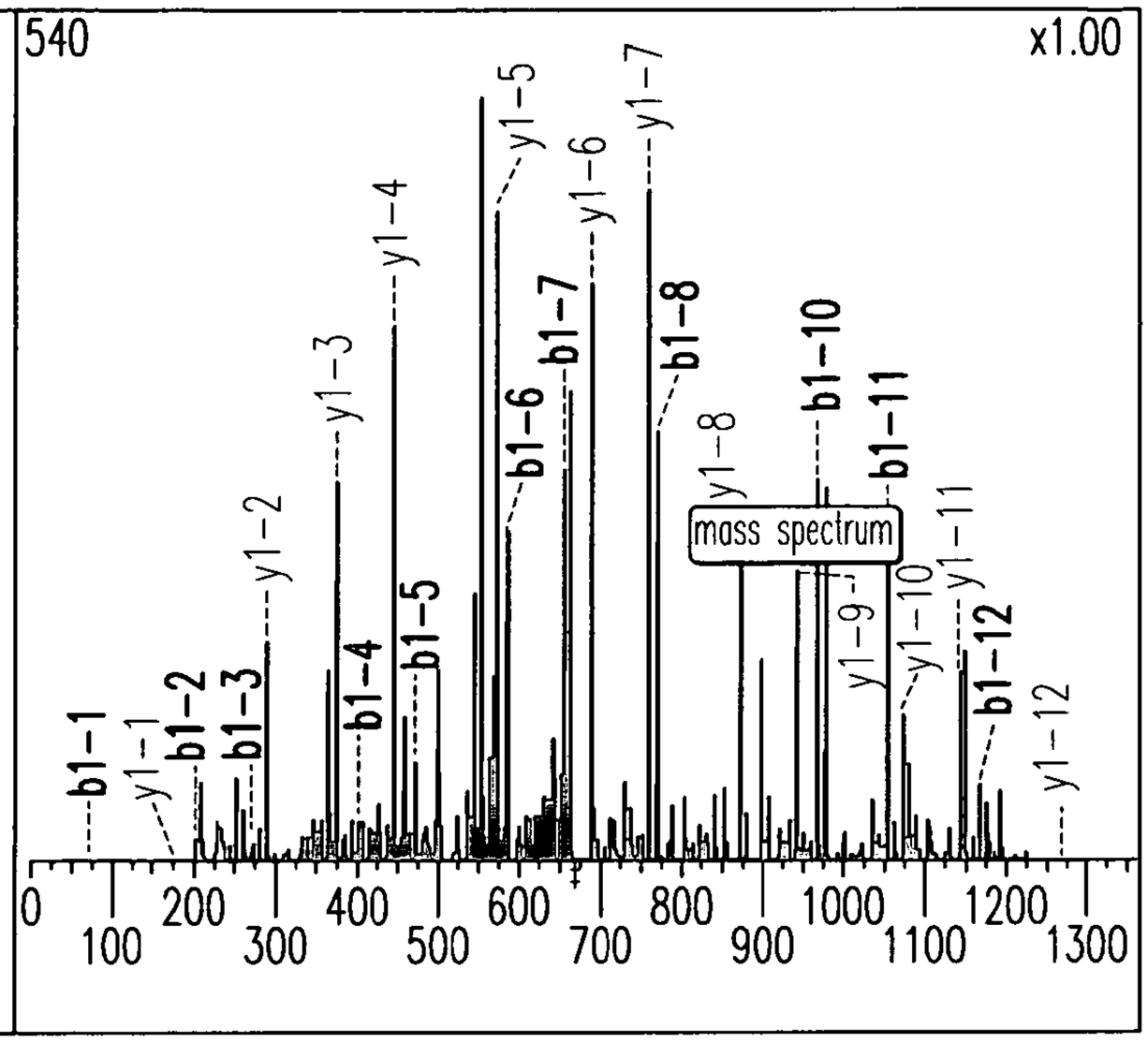


FIG.10E-1

Seq	#	b	y	+1
F	1	148.1	–	6
D	2	**263.1**	613.3	5
P	3	360.1	498.3	4
L	4	473.2	401.3	3
L	5	**586.3**	288.2	2
R	6	–	175.1	1

TACC3 Peptide 2

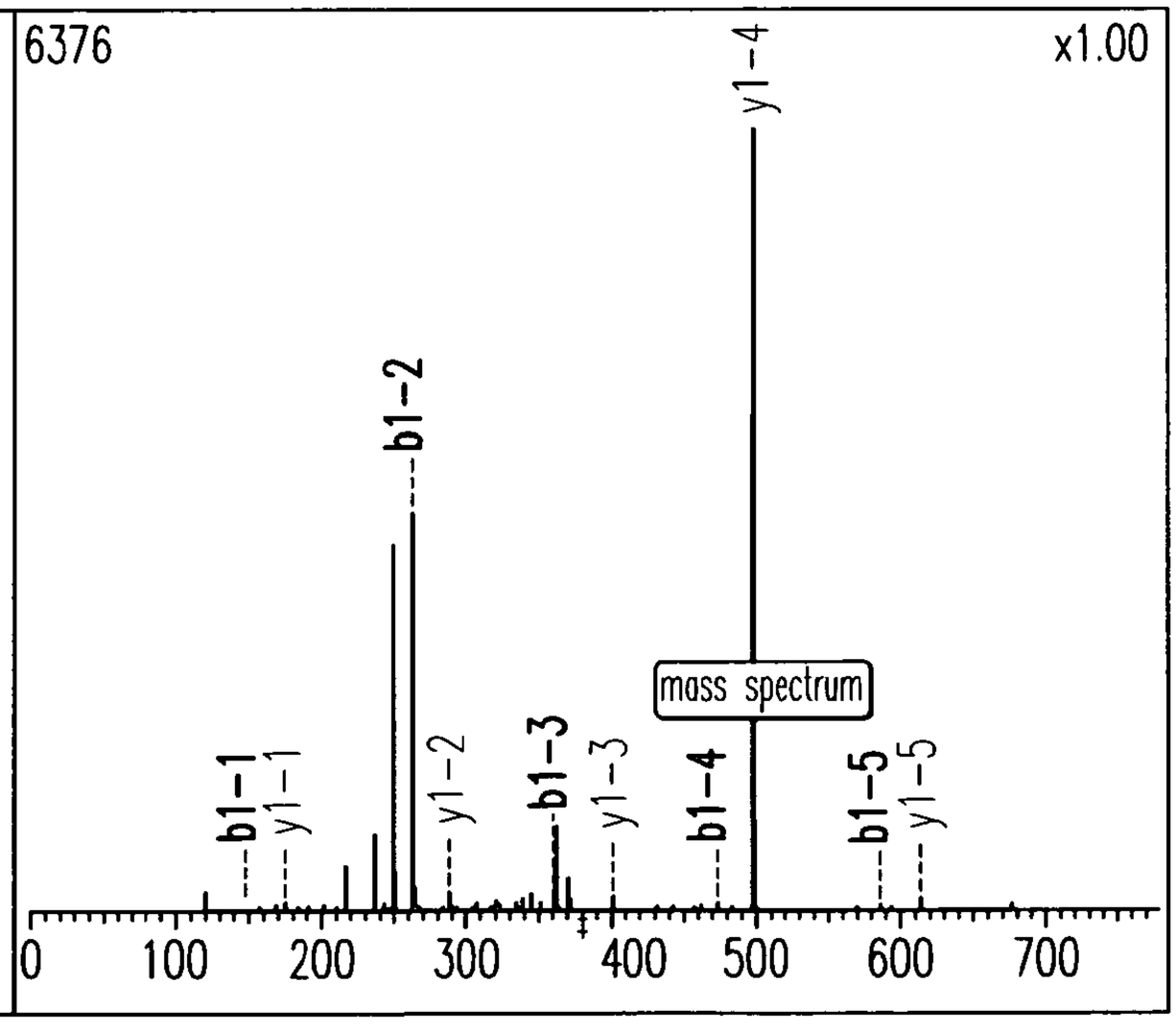


FIG.10E-2

Seq	#	b	y	+1
L	1	114.1	–	12
Q	2	**242.1**	1270.7	11
L	3	**355.2**	1142.6	10
A	4	**426.3**	1029.5	9
N	5	**540.3**	958.5	8
E	6	**669.3**	844.4	7
E	7	798.4	715.4	6
I	8	**911.5**	586.3	5
A	9	**982.5**	473.3	4
Q	10	**1110.6**	402.2	3
V	11	**1209.6**	274.2	2
R	12	–	175.1	1

TACC3
Peptide 3

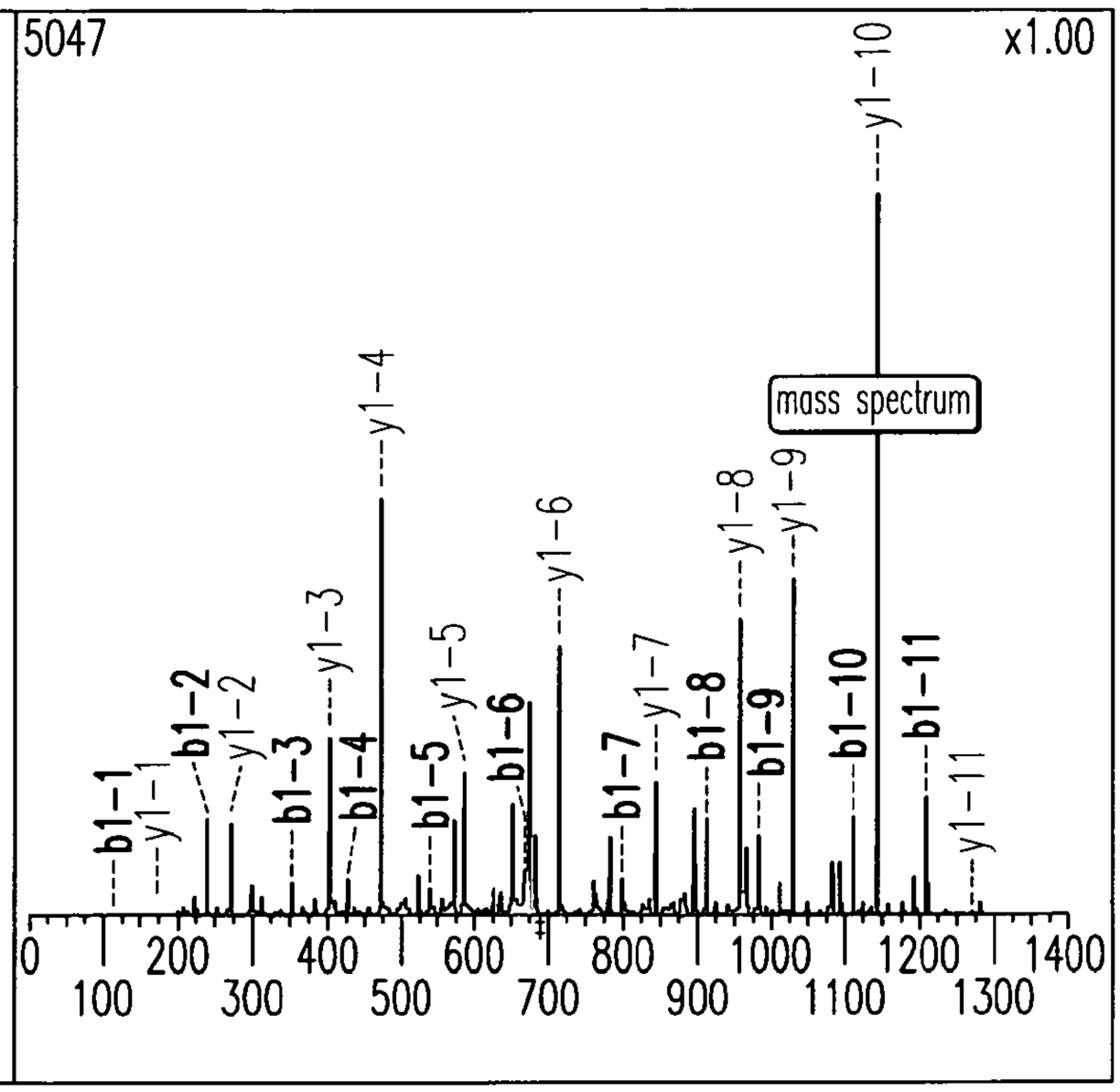


FIG.10E-3

Seq	#	b	y	+1
D	1	116.0	–	10
G	2	**173.0**	915.5	9
T	3	**274.1**	858.5	8
G	4	**331.1**	757.4	7
L	5	**444.2**	700.4	6
V	6	**543.3**	587.3	5
P	7	**640.3**	488.2	4
S	8	727.4	391.2	3
E	9	**856.4**	304.1	2
R	10	–	175.1	1

FGFR3
Peptide 1

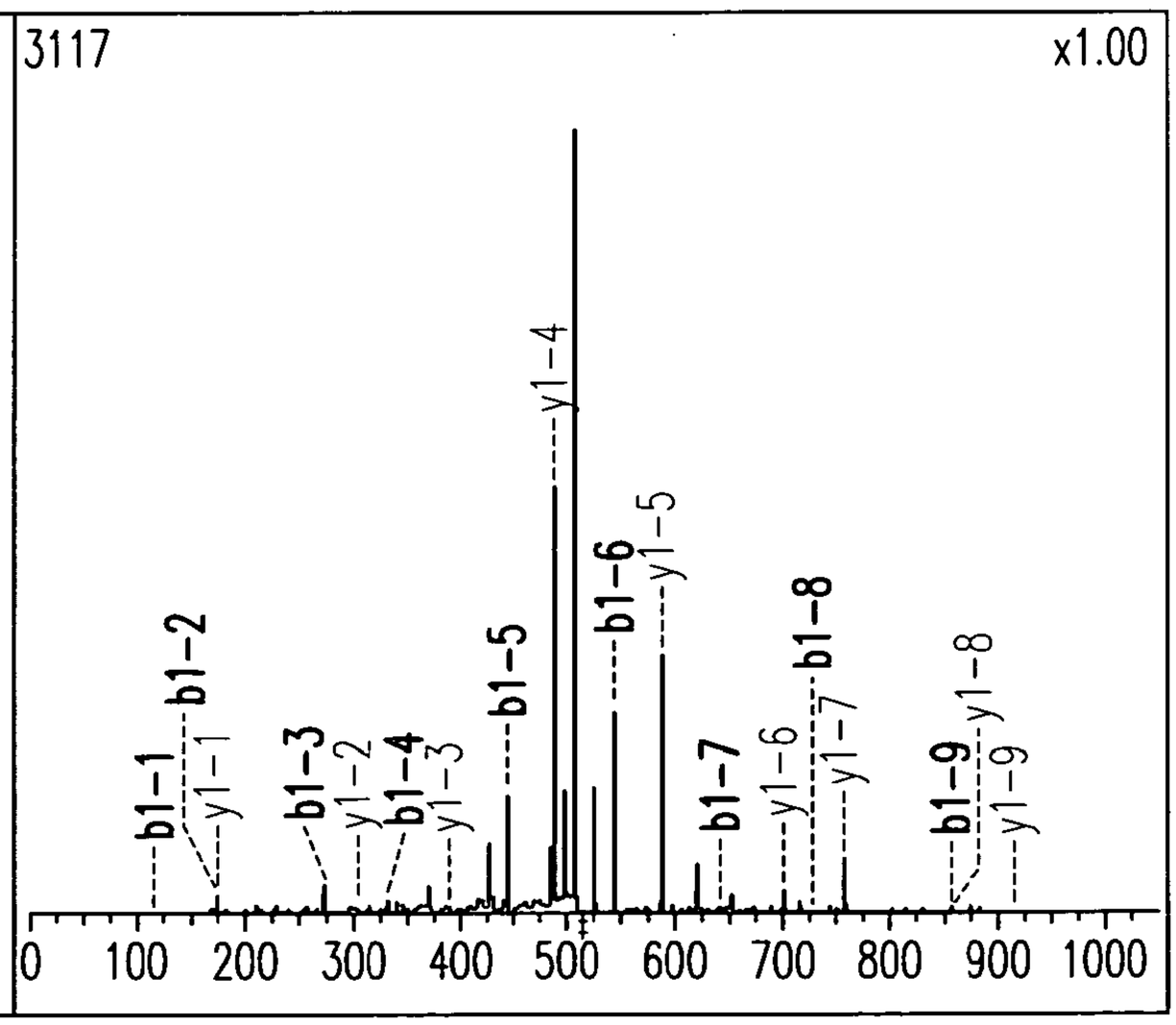


FIG.10E-4

Seq	#	b	y	+1
V	1	100.1	–	13
G	2	157.1	1246.6	12
P	3	**254.1**	1189.6	11
D	4	**369.2**	1092.6	10
G	5	**426.2**	977.5	9
T	6	**527.2**	920.5	8
P	7	**624.3**	819.5	7
Y	8	787.4	722.4	6
V	9	**886.4**	559.4	5
T	10	987.5	460.3	4
V	11	**1086.5**	359.2	3
L	12	**1199.6**	260.2	2
K	13	–	147.1	1

FGFR3
Peptide 2

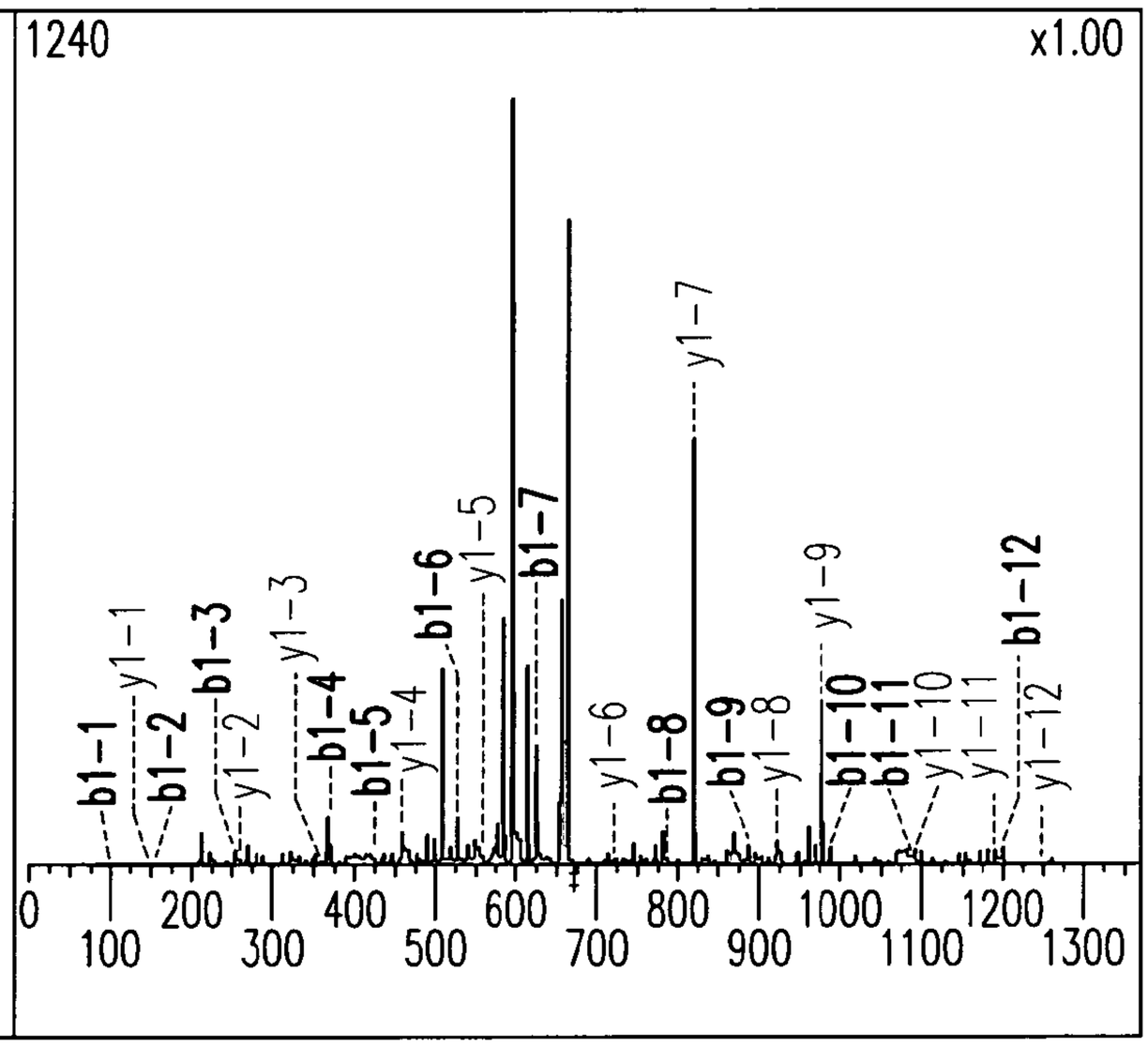


FIG.10E-5

Seq	#	b	y	+1
V	1	100.1	–	7
L	2	**213.2**	669.4	6
V	3	**312.2**	556.3	5
G	4	**369.2**	457.2	4
P	5	466.3	400.2	3
Q	6	**594.4**	303.2	2
R	7	–	175.1	1

FGFR3
Peptide 3

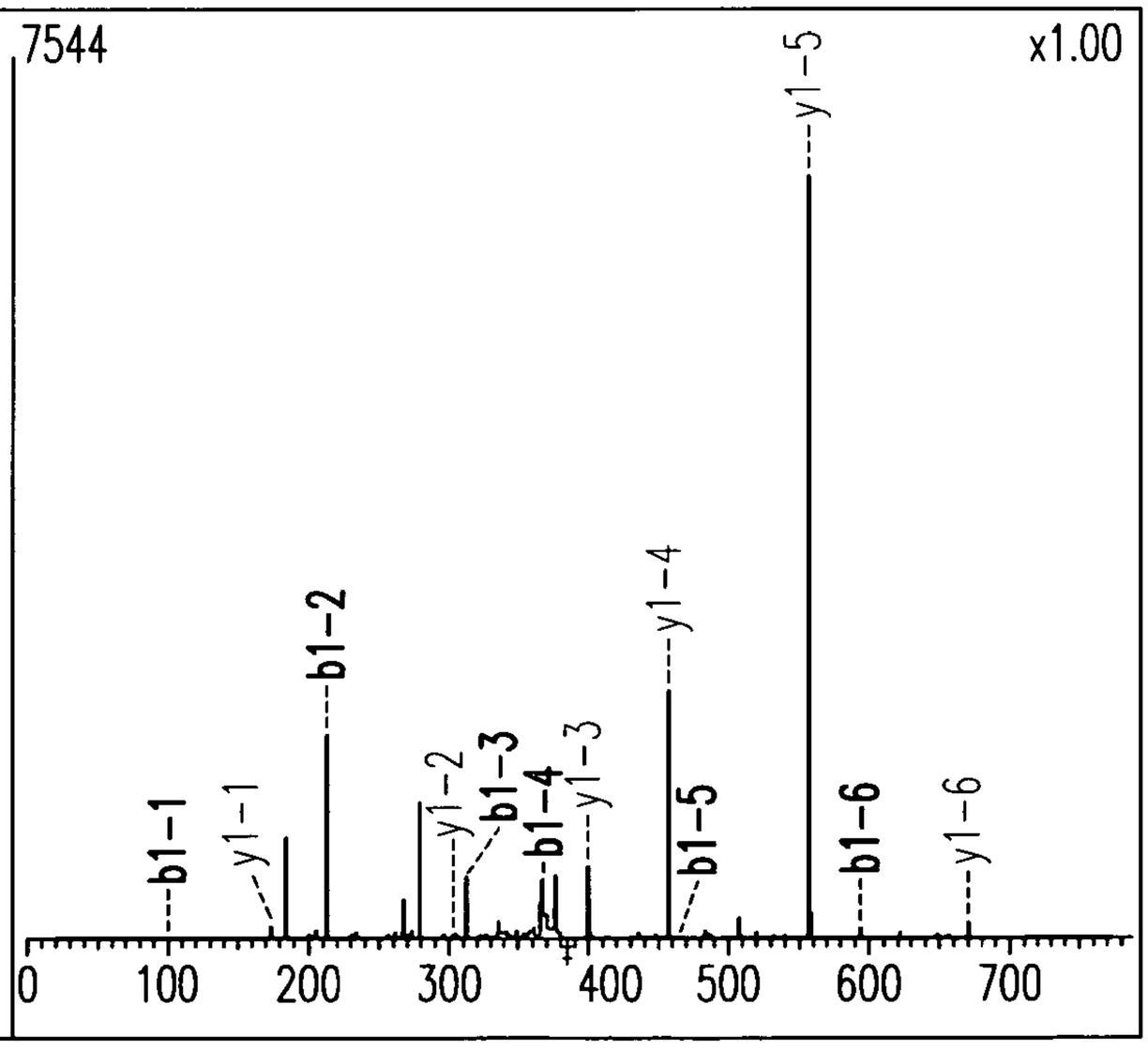
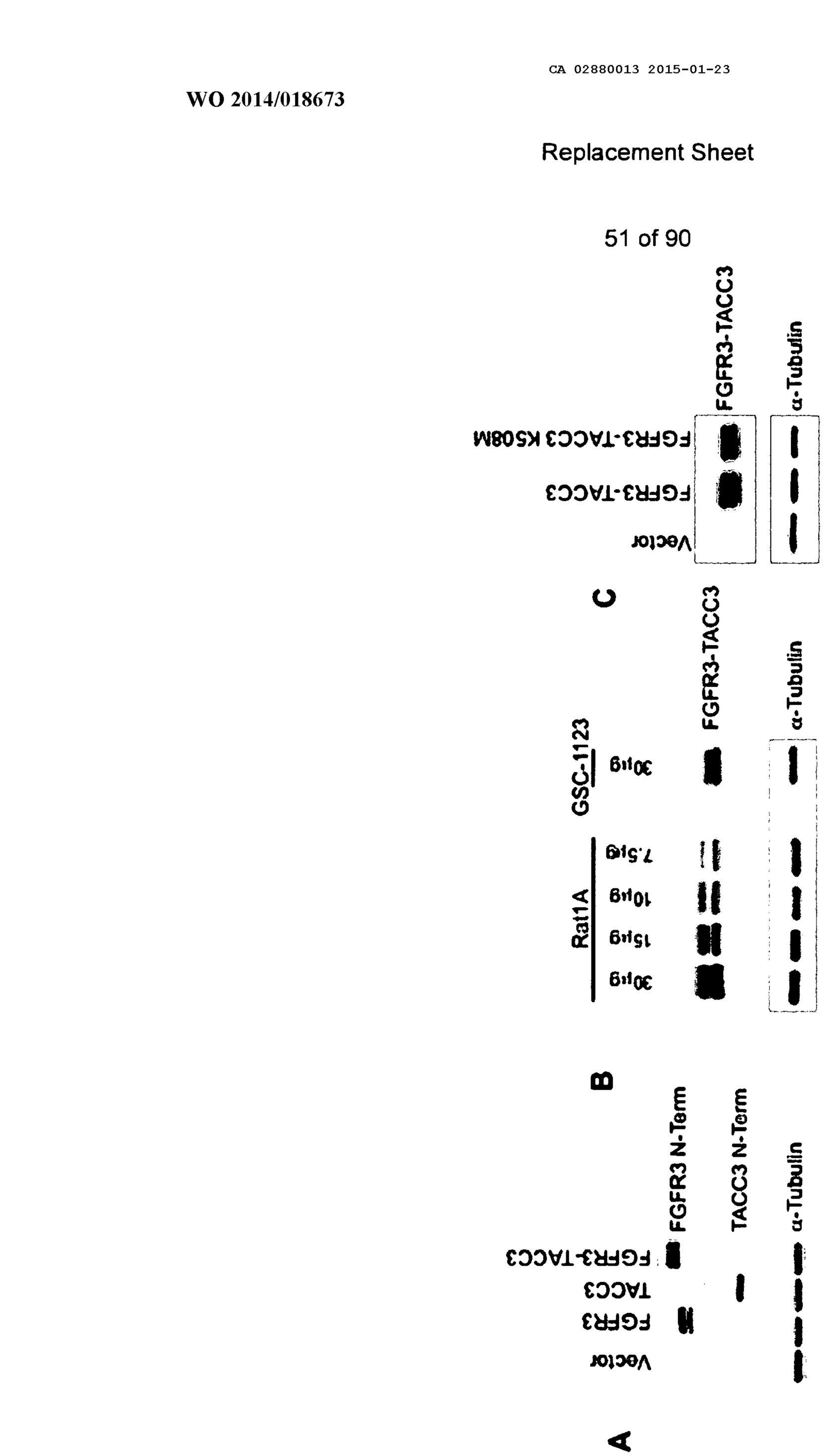
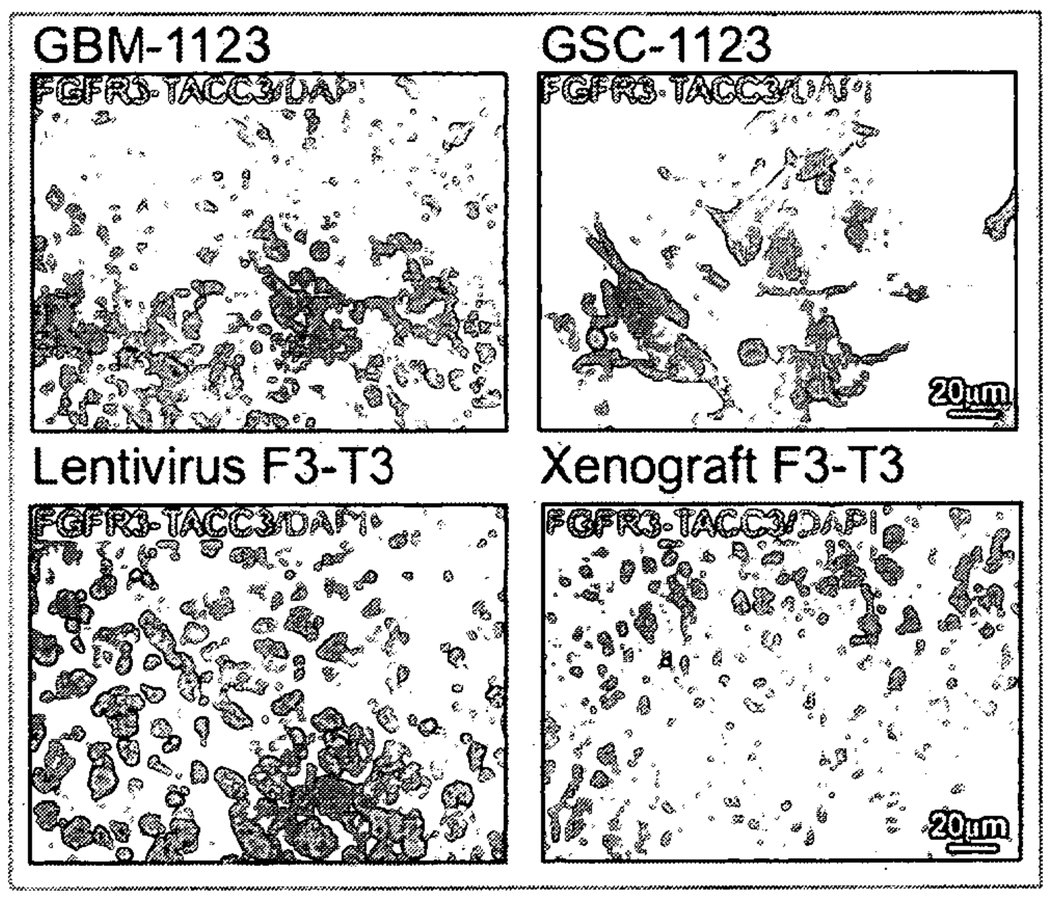


FIG.10E-6

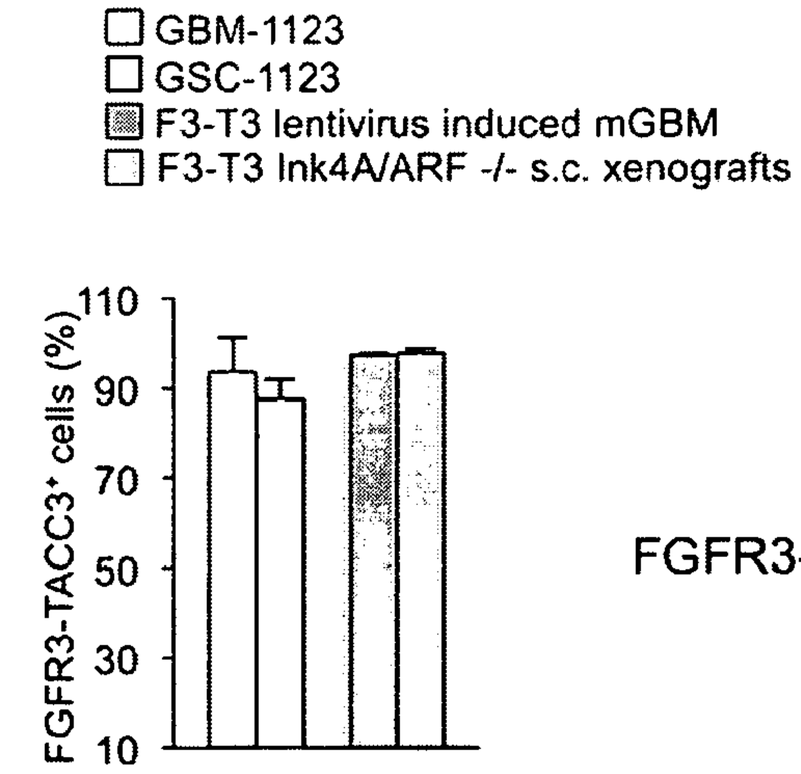

A
Vector
FGFR3-TACC3
TACC3
FGFR3
B
α-Tubulin
TACC3 N-Term
FGFR3 N-Term
FGFR3-TACC3
Rat1A
30μg
15μg
10μg
7.5μg
C
Vector
FGFR3-TACC3
FGFR3-TACC3 K508M
α-Tubulin
GSC-1123
30μg
FGFR3-TACC3
α-Tubulin

FIGS. 11A-C

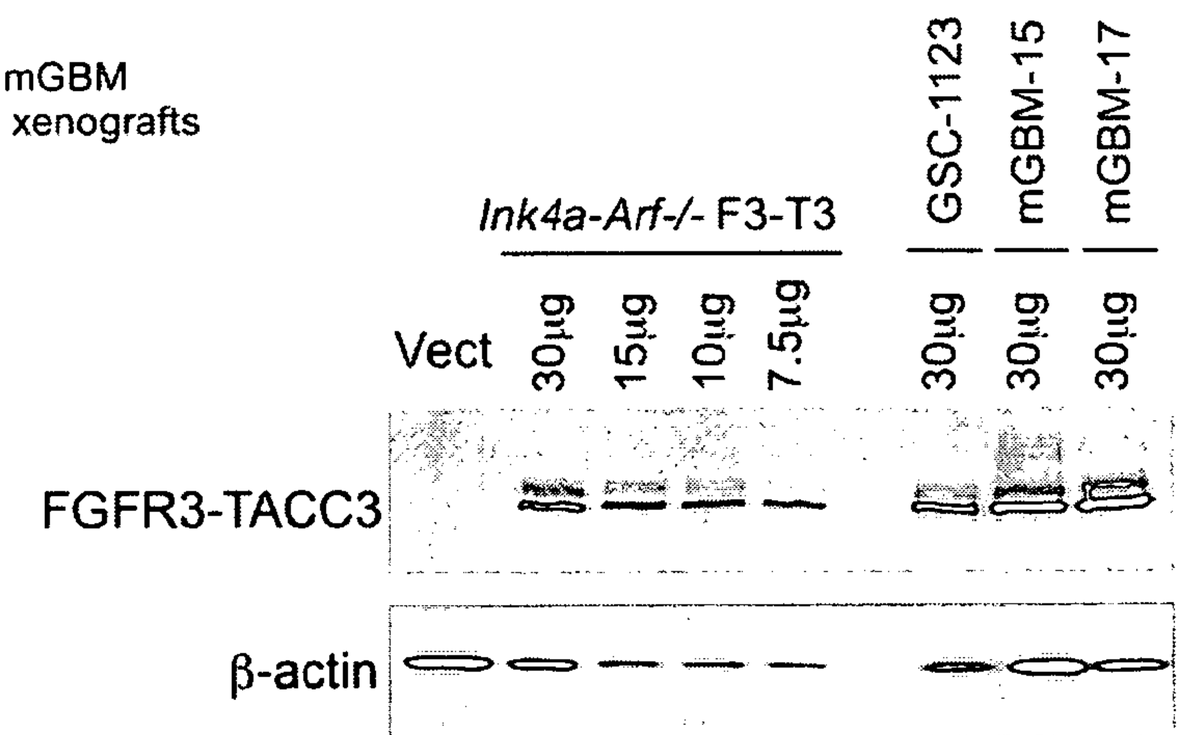
D



E



F



FIGS. 11D-F

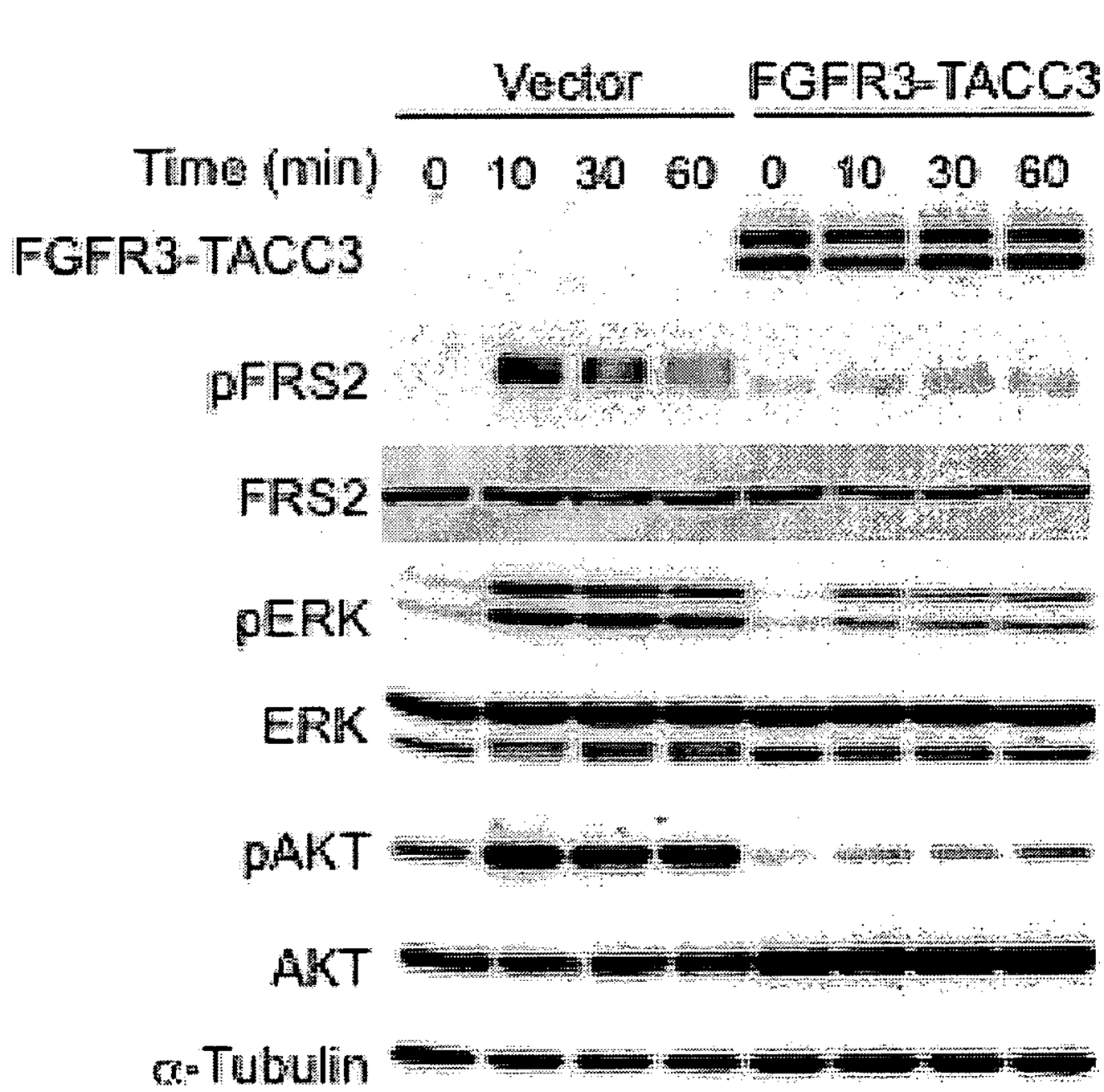

Vector
FGFR3-TACC3
Time (min)
0
10
30
60
0
10
30
60
FGFR3-TACC3
pFRS2
FRS2
pERK
ERK
pAKT
AKT
α-Tubulin

FIG. 12A

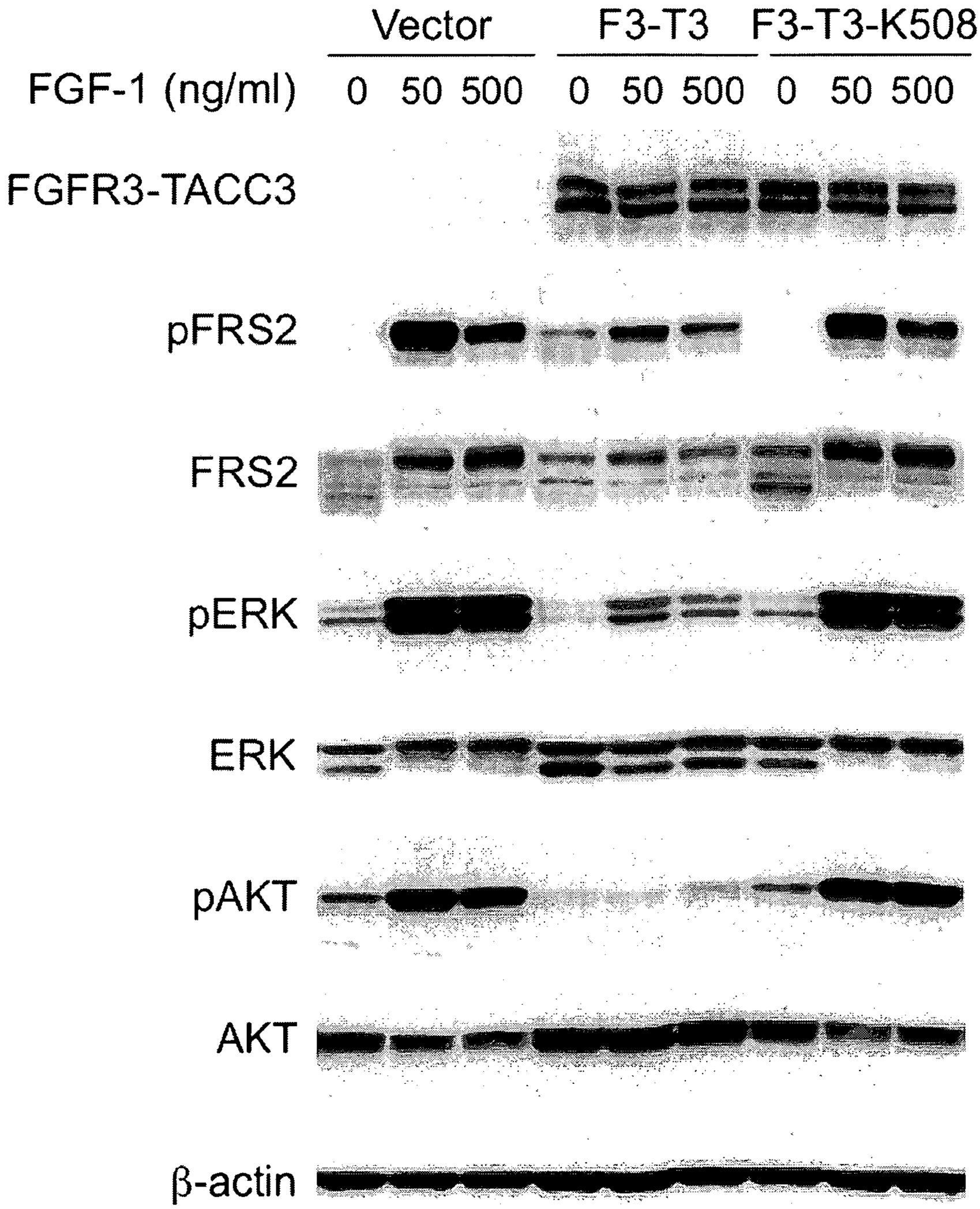

Vector
F3-T3
F3-T3-K508M
FGF-1 (ng/ml)
0 50 500 0 50 500 0 50 500
FGFR3-TACC3
pFRS2
FRS2
pERK
ERK
pAKT
AKT
β-actin

FIG. 12B

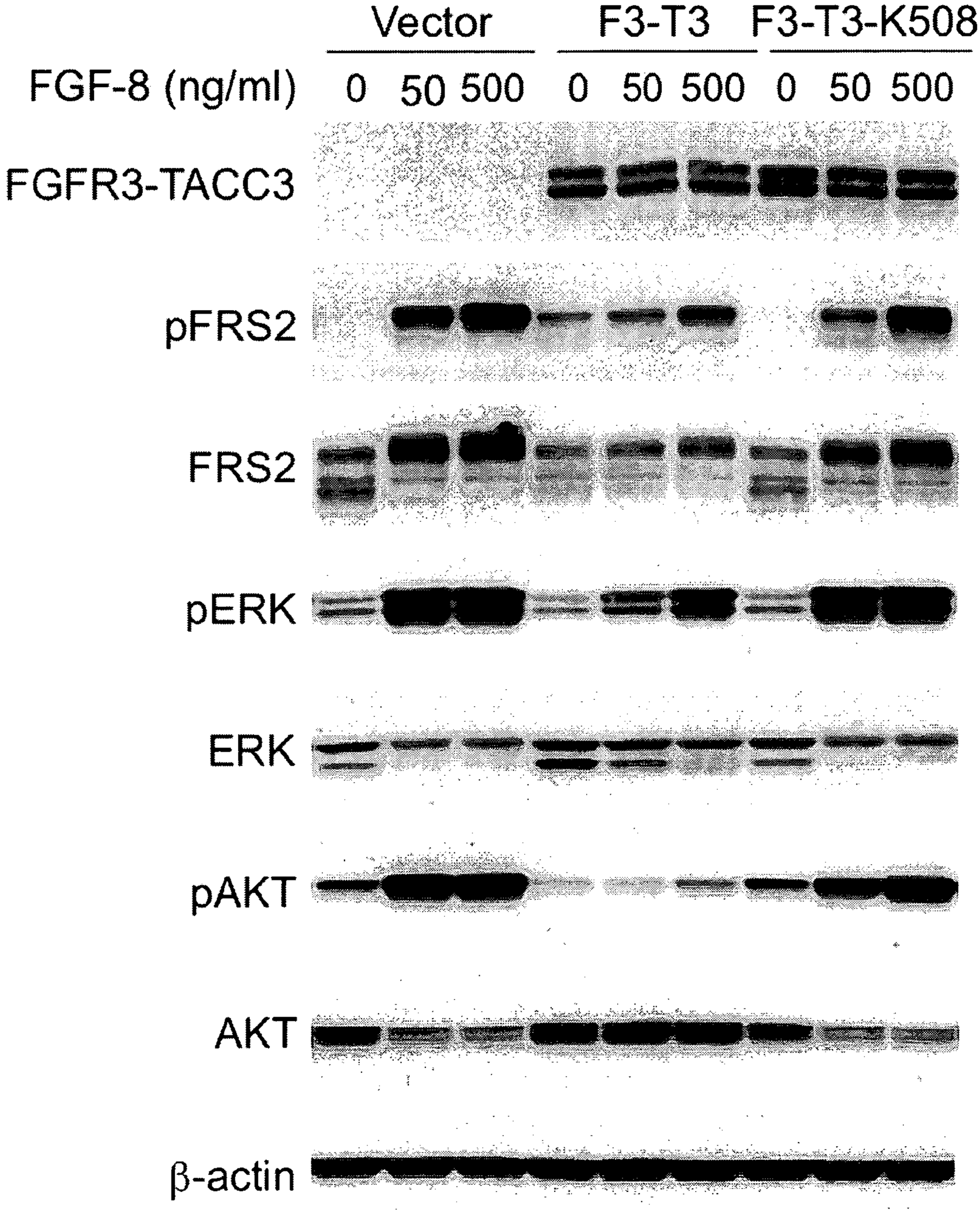
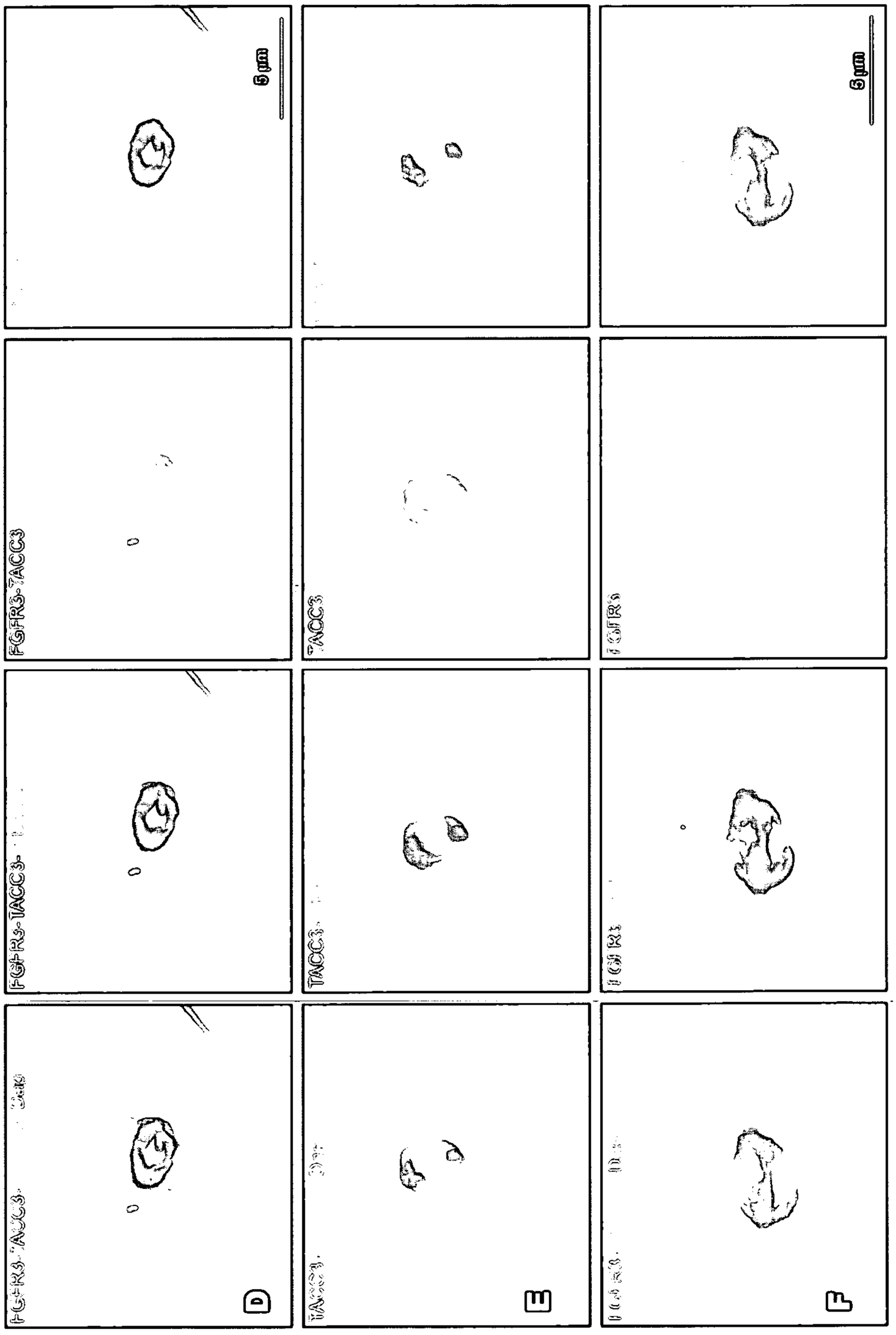

Vector
F3-T3
F3-T3-K508M
FGF-8 (ng/ml)
0 50 500 0 50 500 0 50 500
FGFR3-TACC3
pFRS2
FRS2
pERK
ERK
pAKT
AKT
β-actin

FIG. 12C


FGFR3-TACC3
TACC3
FGFR3
5 μm
6 μm
FGFR3-TACC3
TACC3
FGFR3
FGFR3-TACC3
TACC3
FGFR3
D
E
F

FIGS. 12D-F

Cell line	Metaphases inspected	Cells with segregation defects (%±SD)	Anaphases with lagging chromosomes (%±SD)	Anaphases/telophases with chromosome bridges (%±SD)
Ink4A;ARF-/- Vector	100	7.3±1.5	2.0±0.7	3.3±0.6
Ink4A;ARF-/- F3-T3 Fusion	100	18.3±1.5	6.5±1.3	11.2±0.7

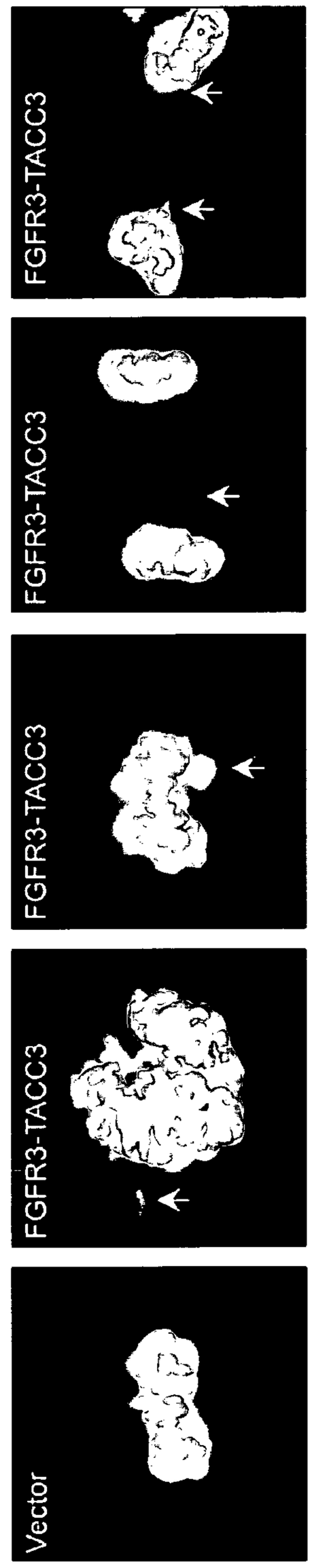
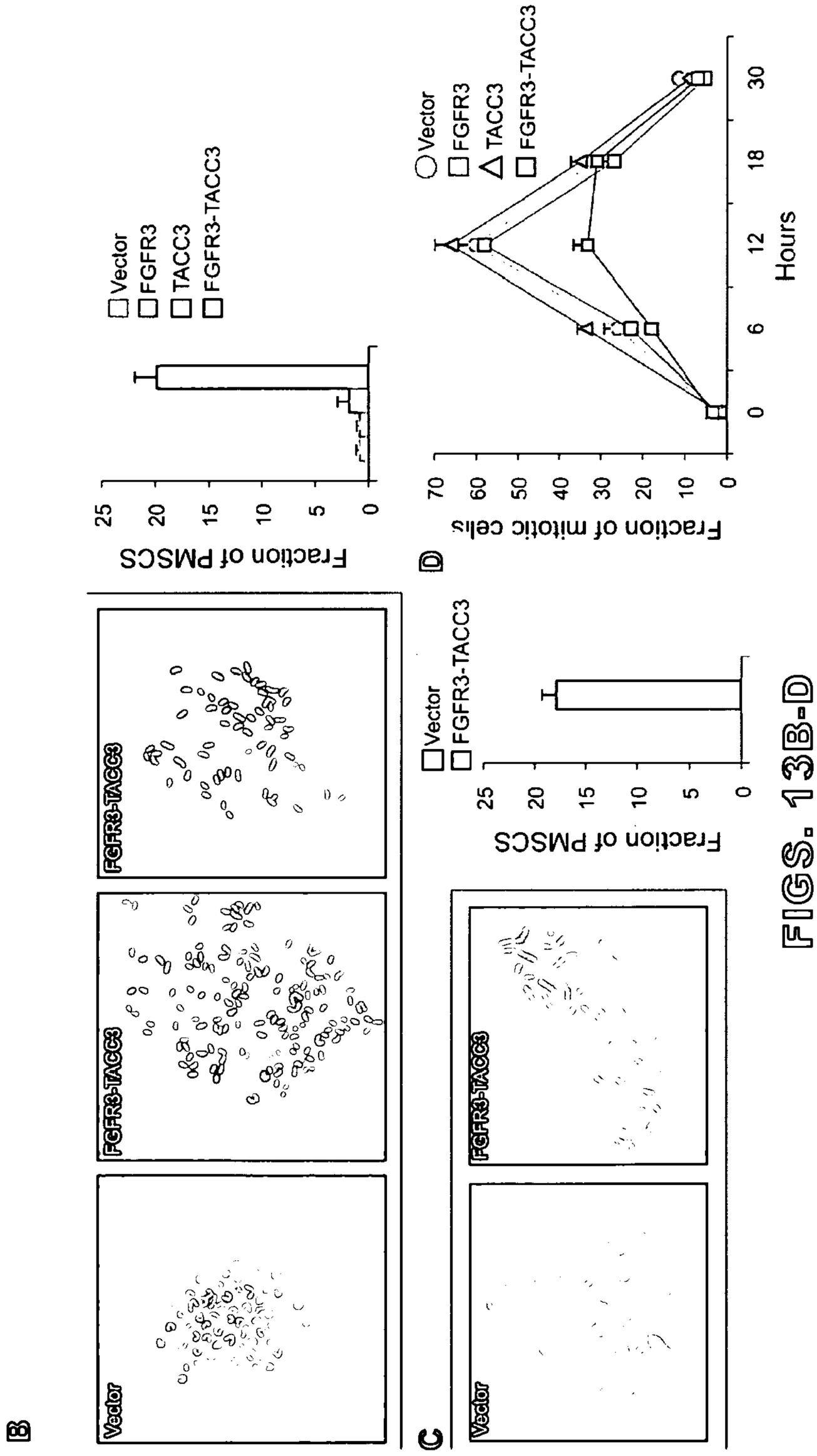


FIG. 13A


B
Vector
FGFR3-TACC3
C
Vector
FGFR3-TACC3
D
Vector
FGFR3
TACC3
FGFR3-TACC3
Fraction of mitotic cells
Hours
0
6
12
18
30
10
20
30
40
50
60
70
Fraction of PMSCS
0
5
10
15
20
25
Vector
FGFR3
TACC3
FGFR3-TACC3
Fraction of PMSCS
0
5
10
15
20
25
Vector
FGFR3-TACC3

FIGS. 13B-D

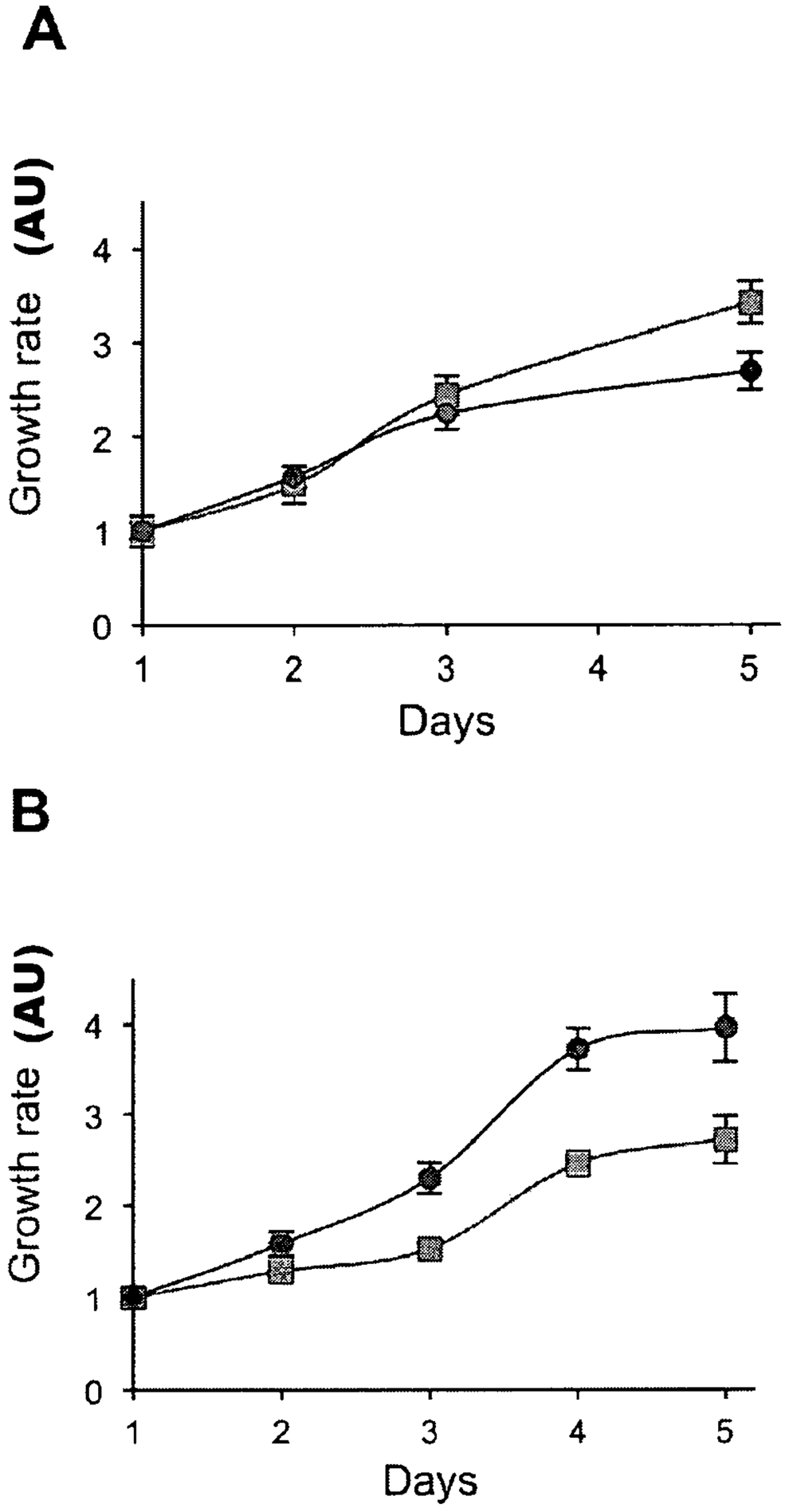
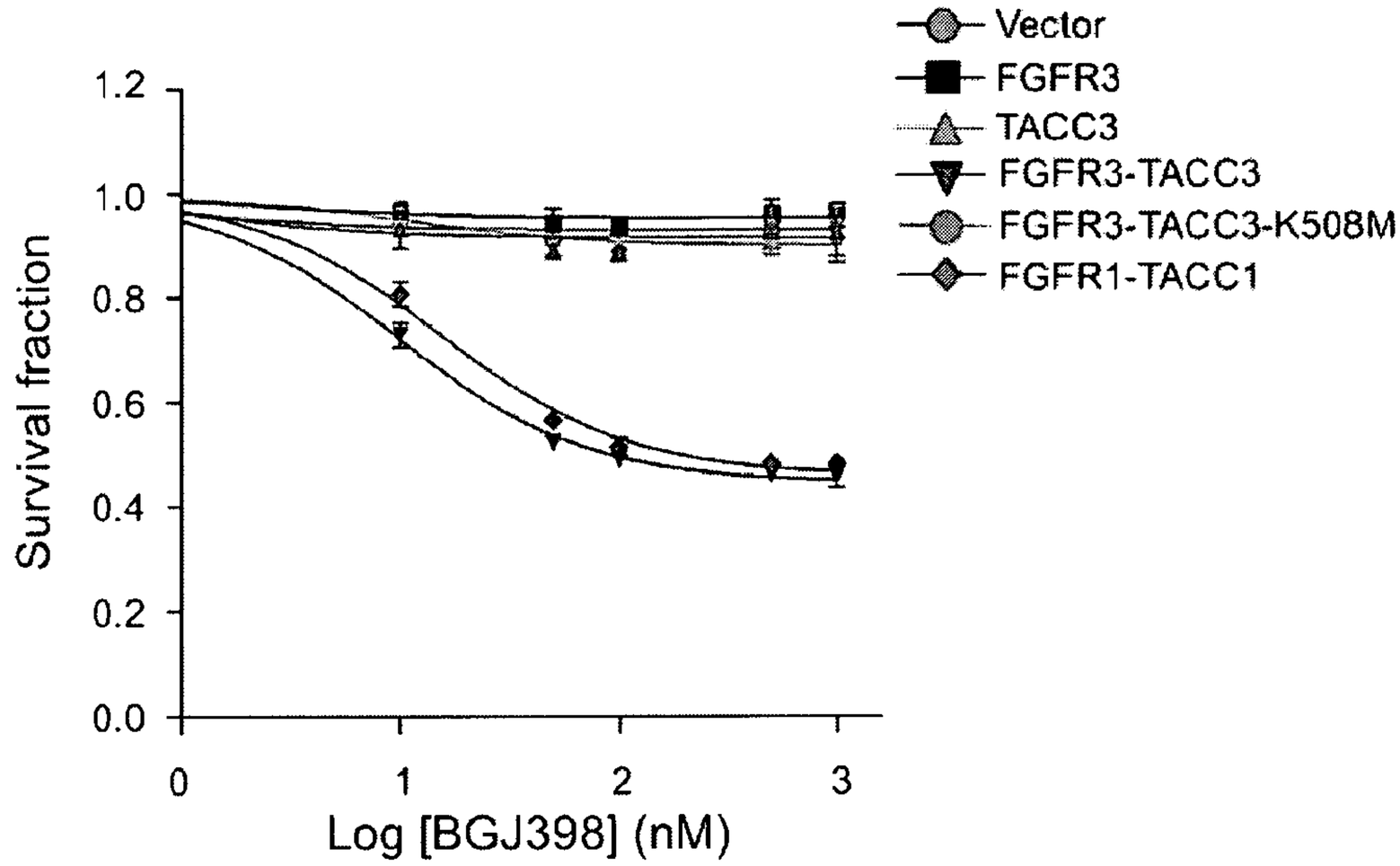

A
Vector
FGFR3-TACC3
Growth rate (AU)
0
1
2
3
4
1
2
3
4
5
Days
B
Growth rate (AU)
0
1
2
3
4
1
2
3
4
5
Days

FIG. 14A-B

C


Vector
FGFR3
TACC3
FGFR3-TACC3
FGFR3-TACC3-K508M
FGFR1-TACC1
Survival fraction
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
Log [BGJ398] (nM)

D

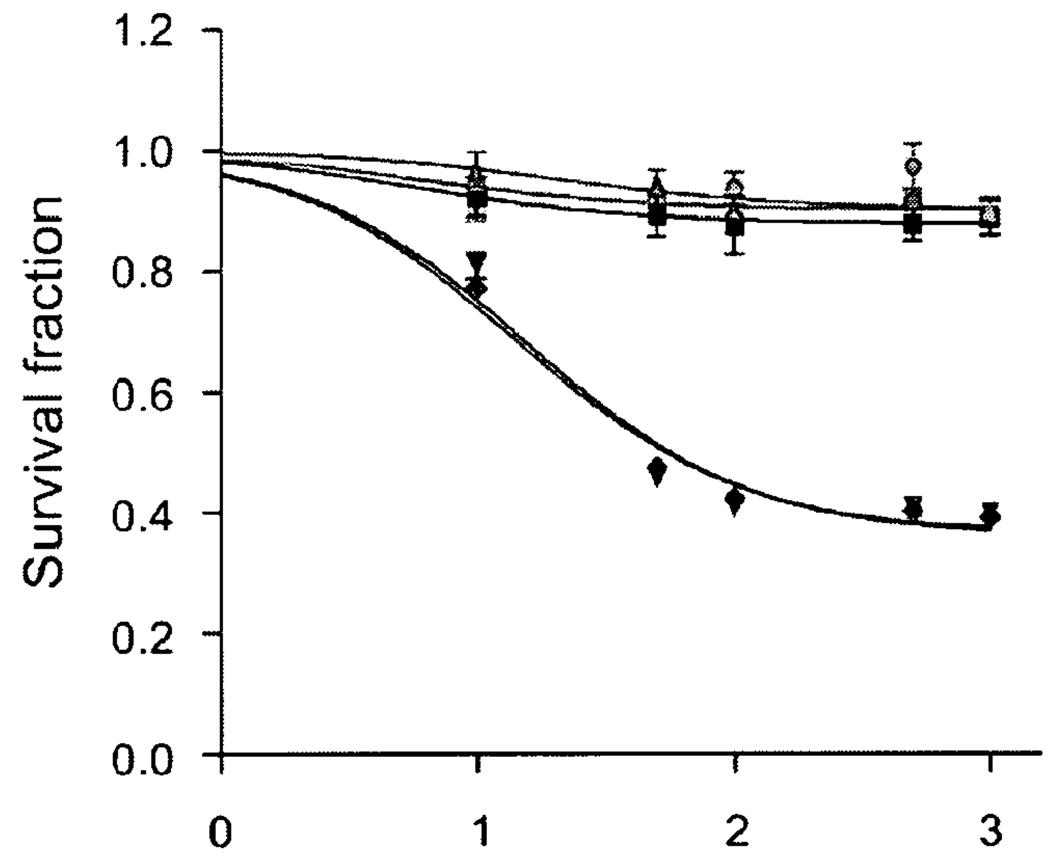

Survival fraction
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
Log [AZD4547] (nM)

FIG. 14C-D

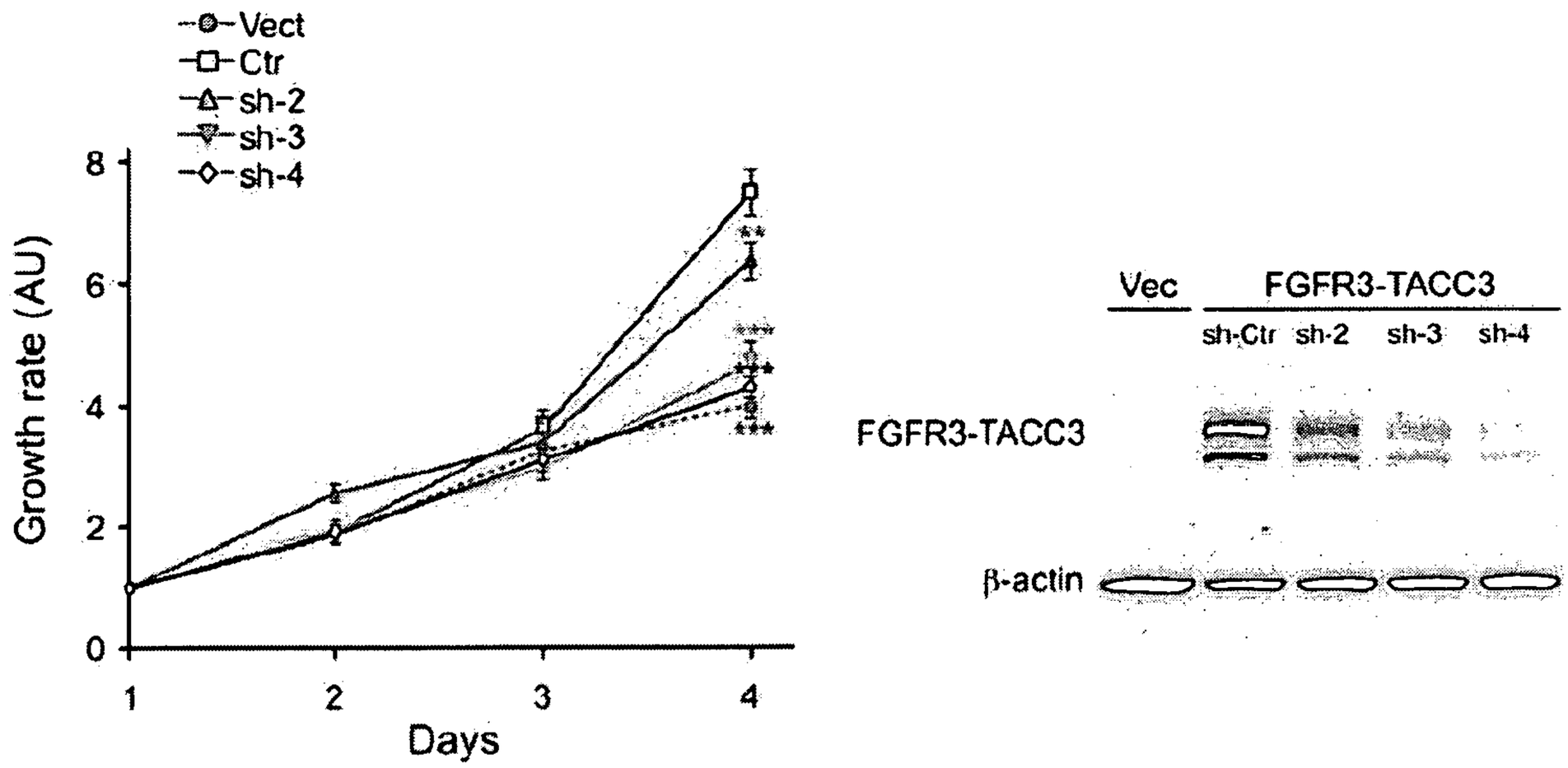


FIG.14E

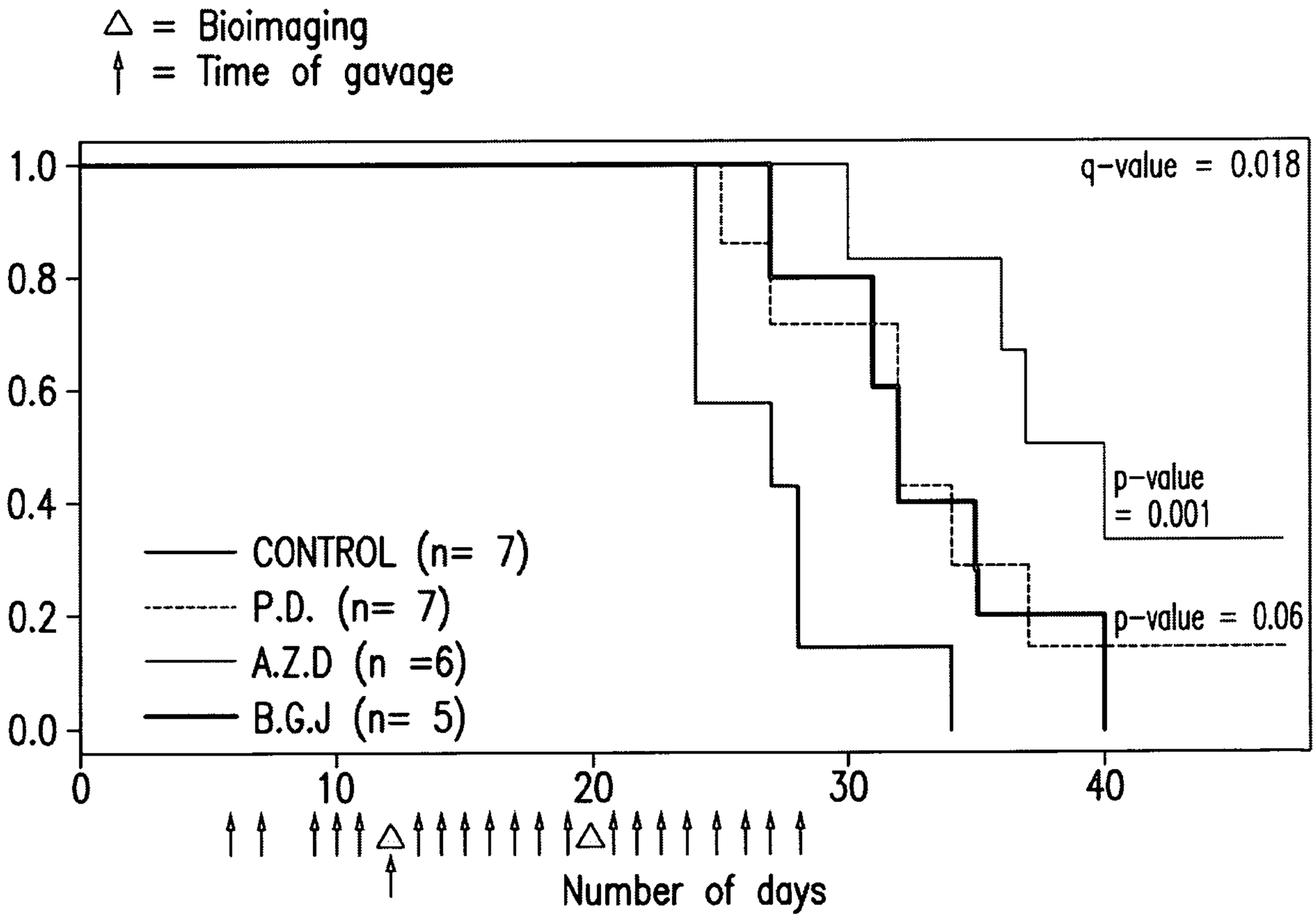


FIG.15

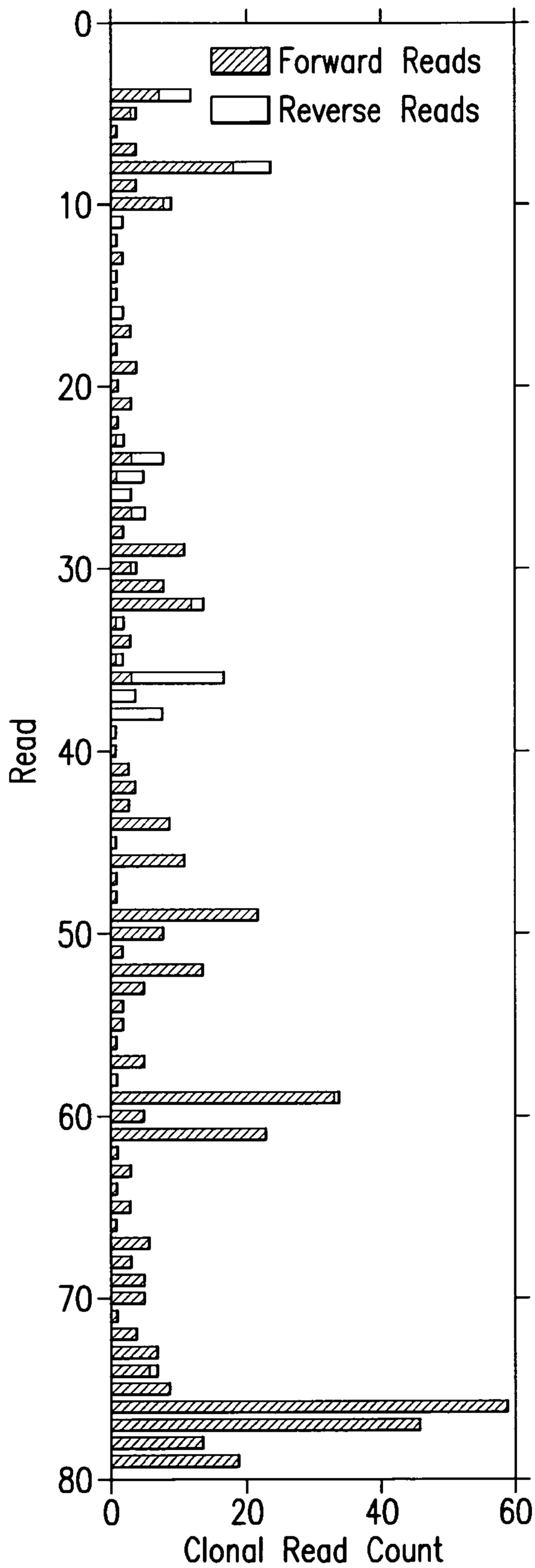

Forward Reads
Reverse Reads
Read
0
10
20
30
40
50
60
70
80
0
20
40
60
Clonal Read Count

FIG.16

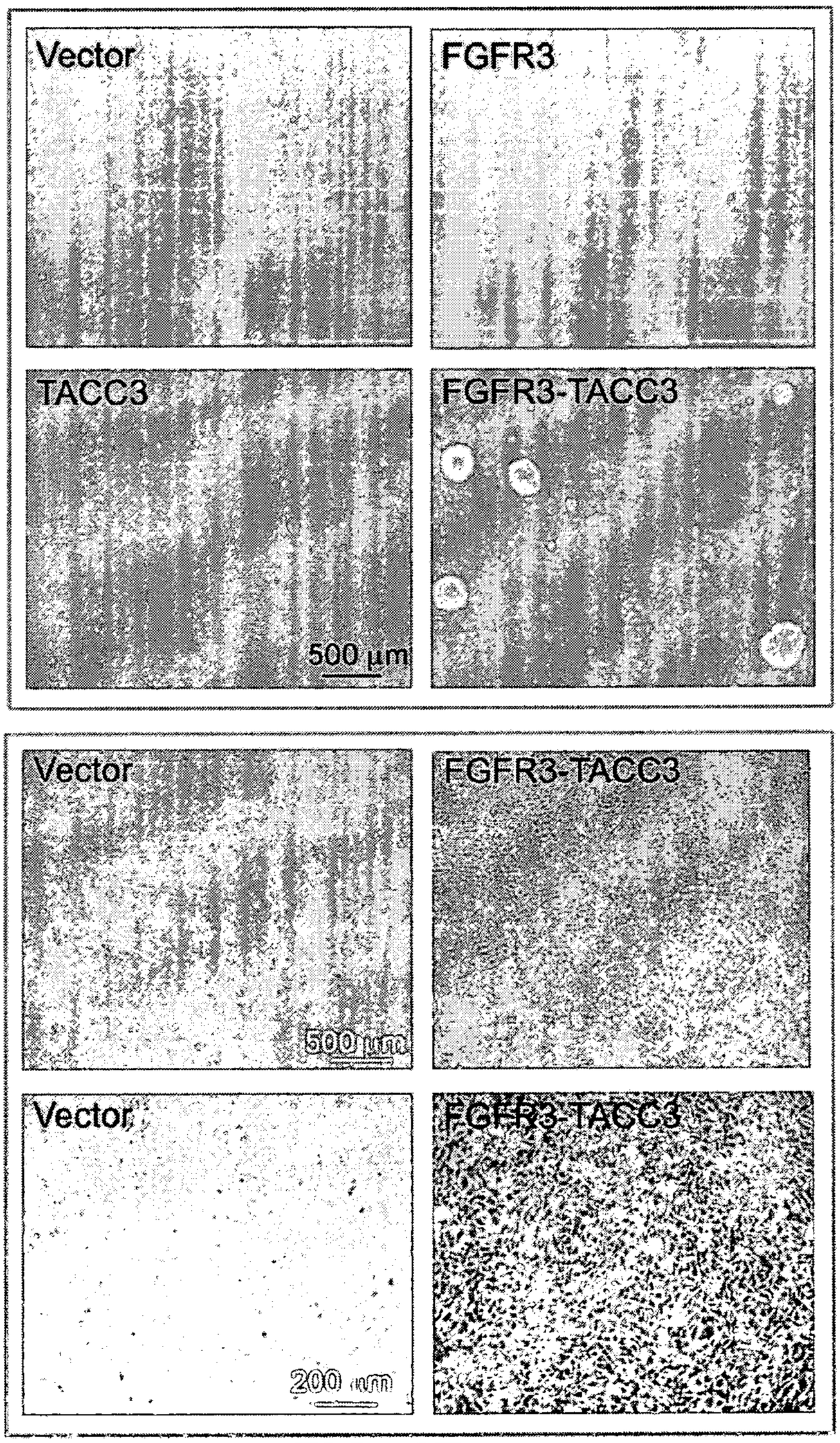

Vector
FGFR3
TACC3
500 µm
FGFR3-TACC3
Vector
FGFR3-TACC3
500 µm
Vector
FGFR3-TACC3
200 µm

FIG. 17

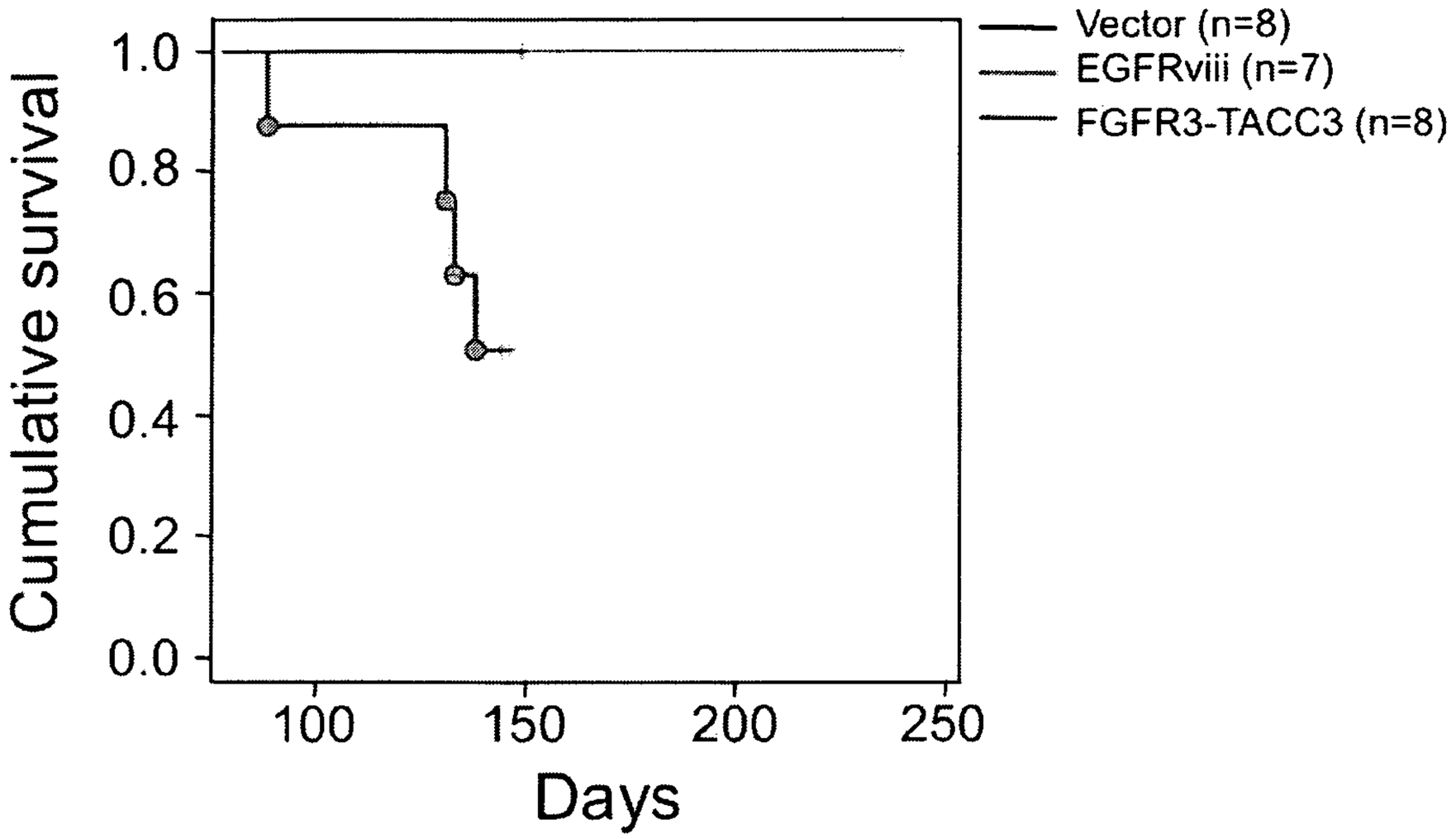

Vector (n=8)
EGFRviii (n=7)
FGFR3-TACC3 (n=8)
Cumulative survival
1.0
0.8
0.6
0.4
0.2
0.0
100
150
200
250
Days

FIG. 18

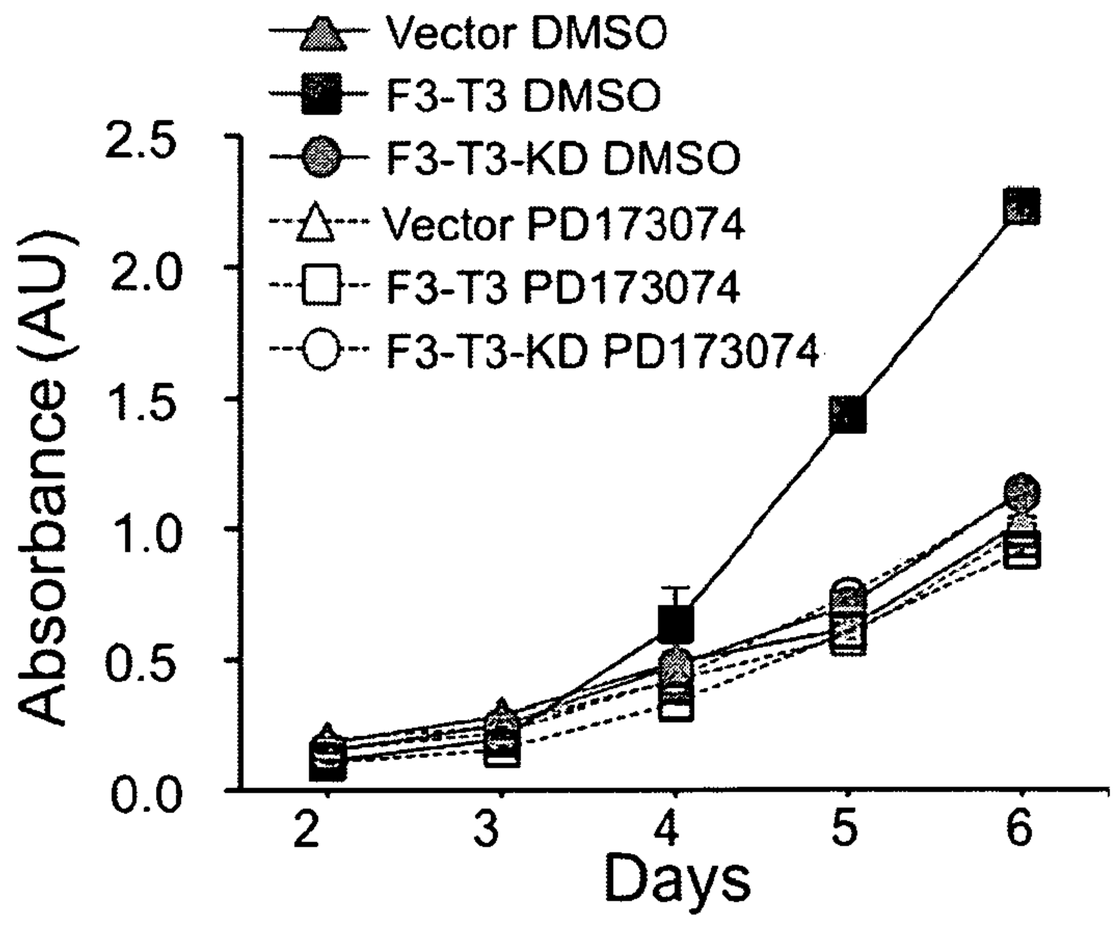

Vector DMSO
F3-T3 DMSO
F3-T3-KD DMSO
Vector PD173074
F3-T3 PD173074
F3-T3-KD PD173074
Absorbance (AU)
2.5
2.0
1.5
1.0
0.5
0.0
2
3
4
5
6
Days

FIG. 19

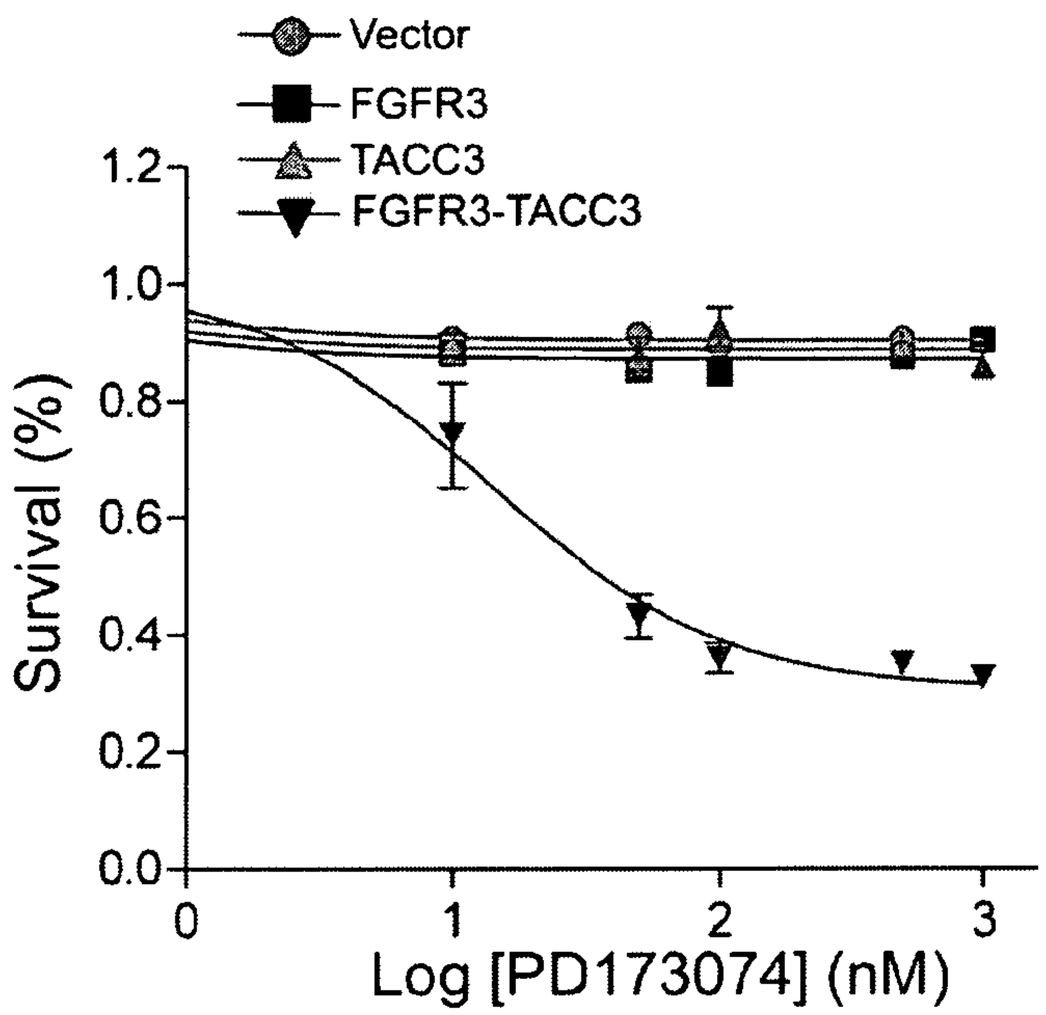

Vector
FGFR3
TACC3
FGFR3-TACC3
Survival (%)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
Log [PD173074] (nM)

FIG. 20

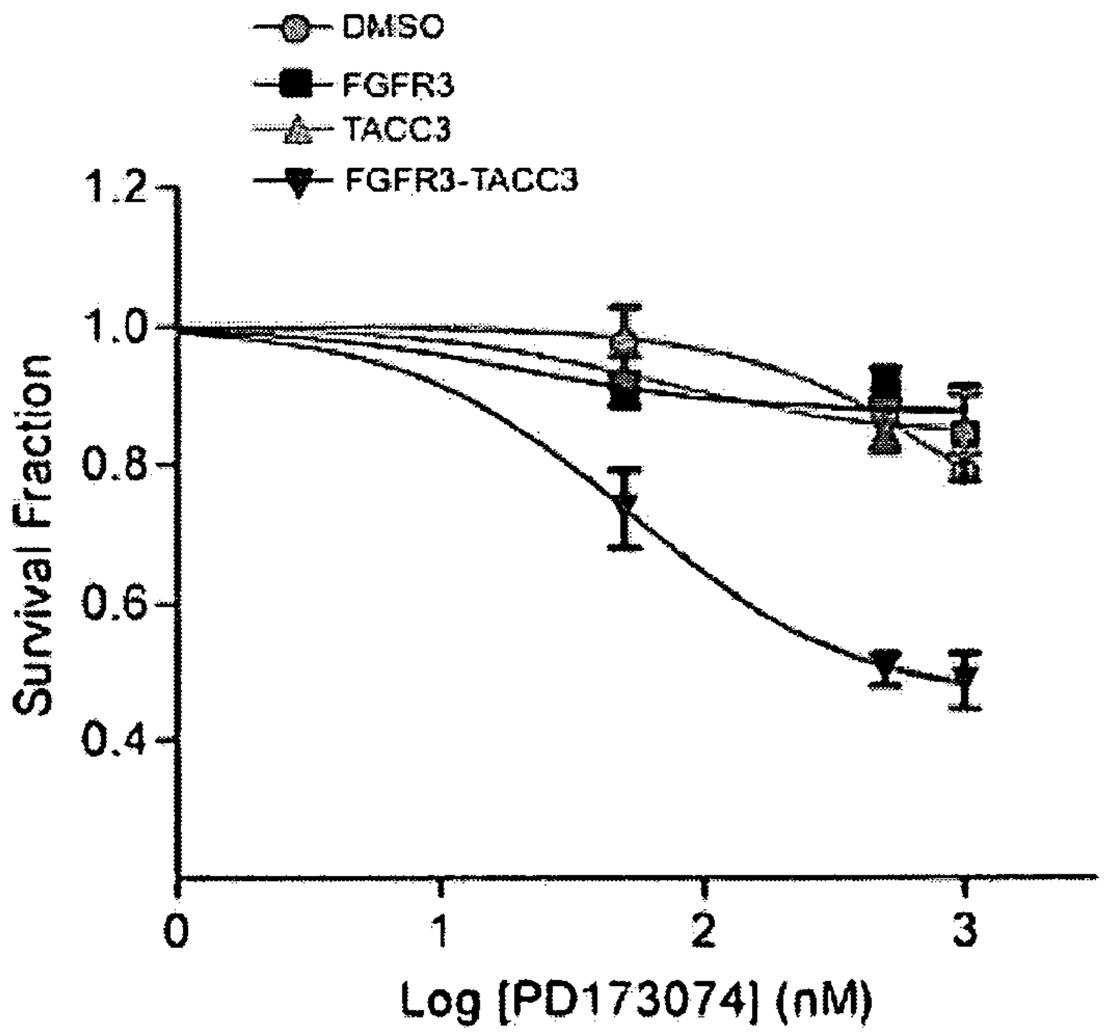
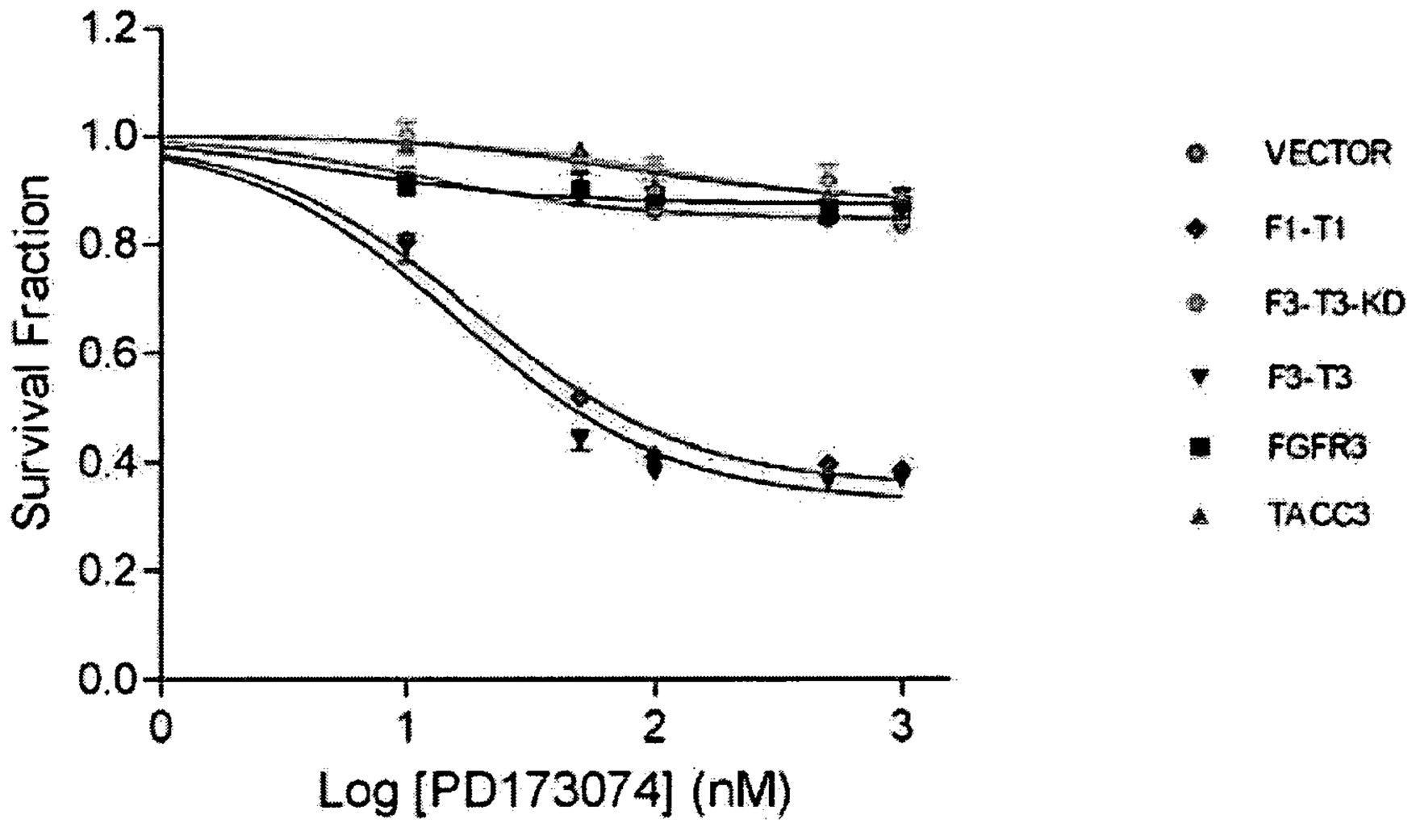

DMSO
FGFR3
TACC3
FGFR3-TACC3
Survival Fraction
1.2
1.0
0.8
0.6
0.4
0
1
2
3
Log [PD173074] (nM)

FIG. 21


1.2
1.0
0.8
0.6
0.4
0.2
0.0
Survival Fraction
0
1
2
3
Log [PD173074] (nM)
VECTOR
F1-T1
F3-T3-KD
F3-T3
FGFR3
TACC3

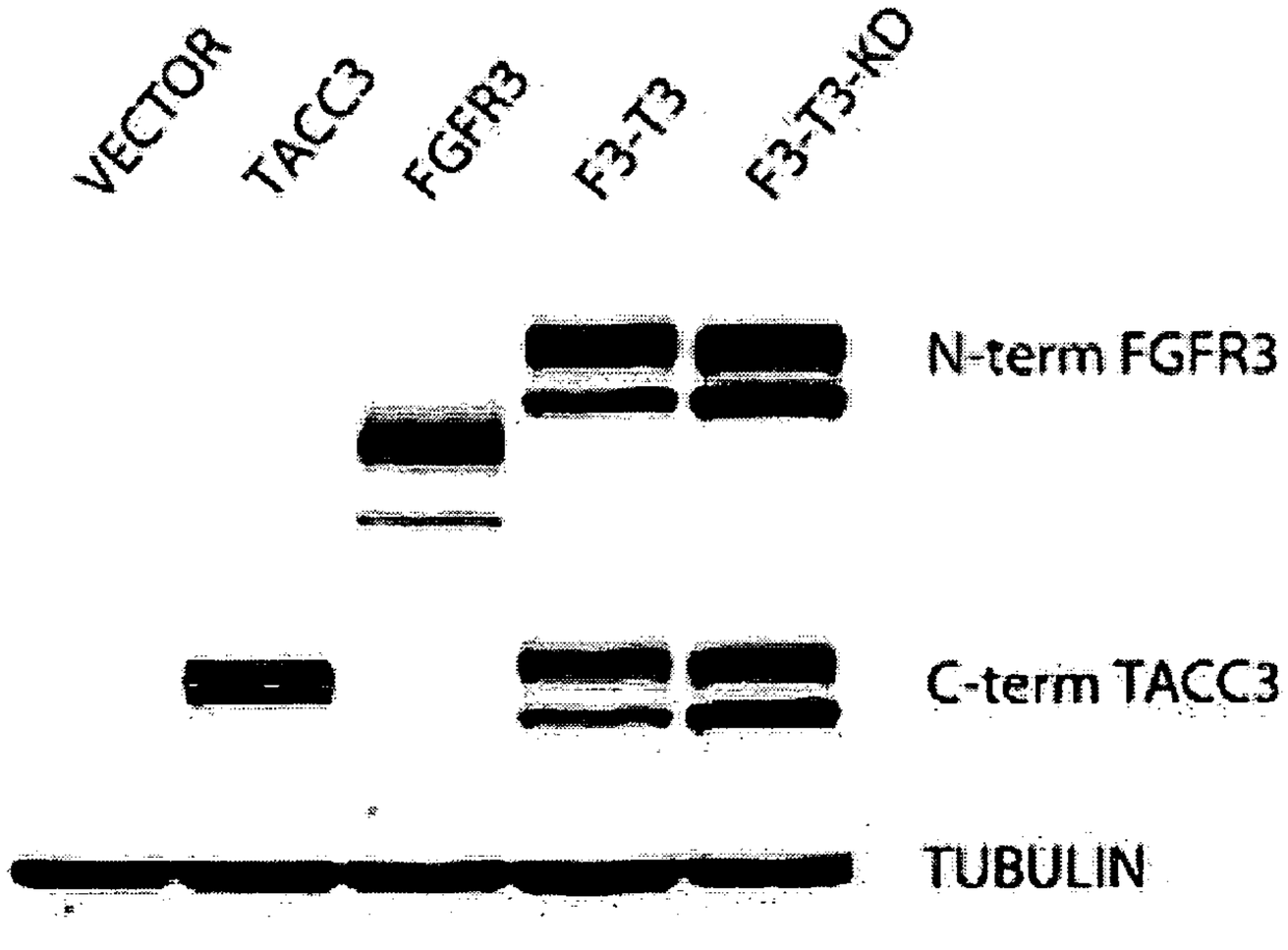

VECTOR
TACC3
FGFR3
F3-T3
F3-T3-KD
N-term FGFR3
C-term TACC3
TUBULIN

FIG. 22

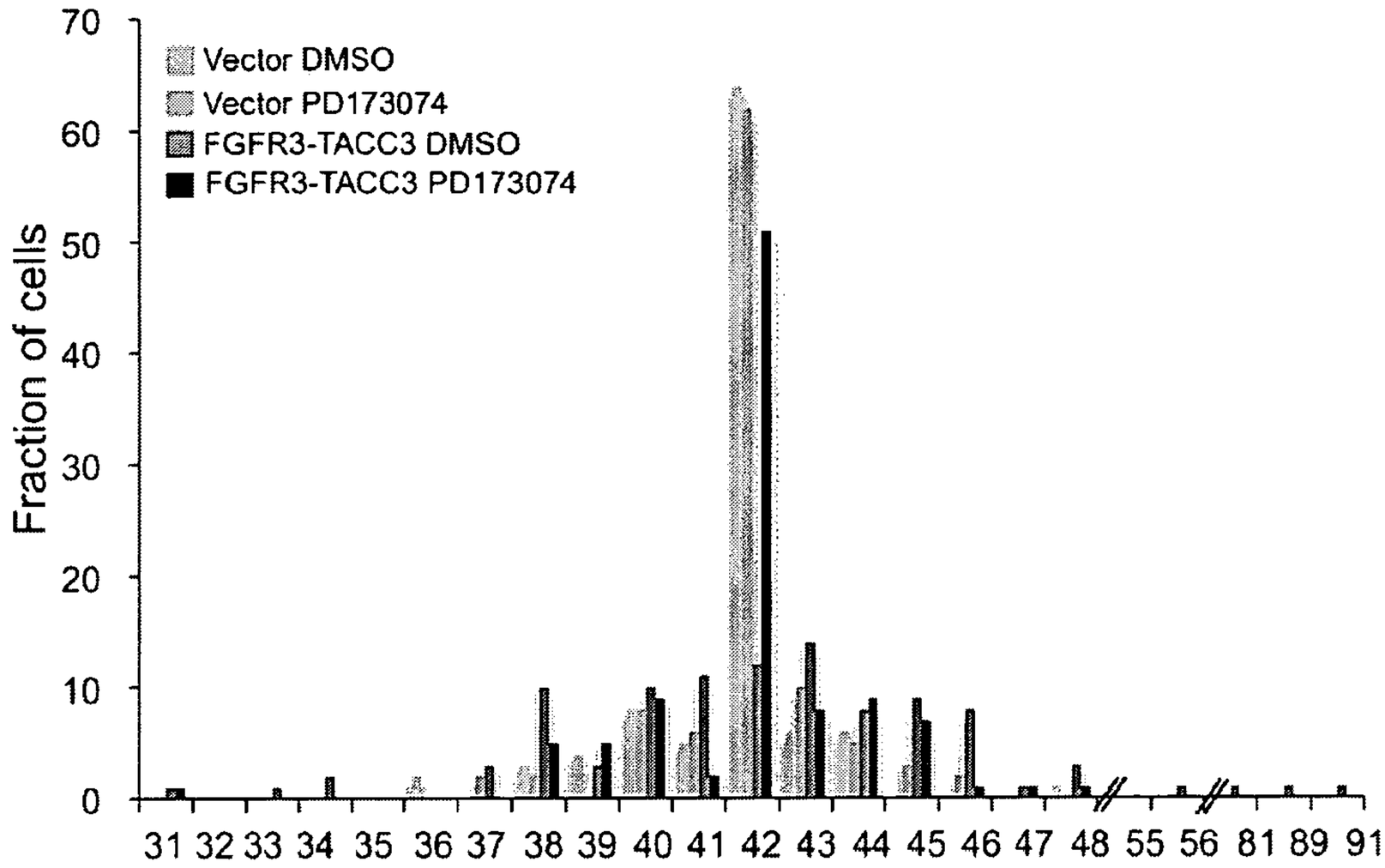

Vector DMSO
Vector PD173074
FGFR3-TACC3 DMSO
FGFR3-TACC3 PD173074
Fraction of cells
70
60
50
40
30
20
10
0
31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 55 56 81 89 91

FIG. 23

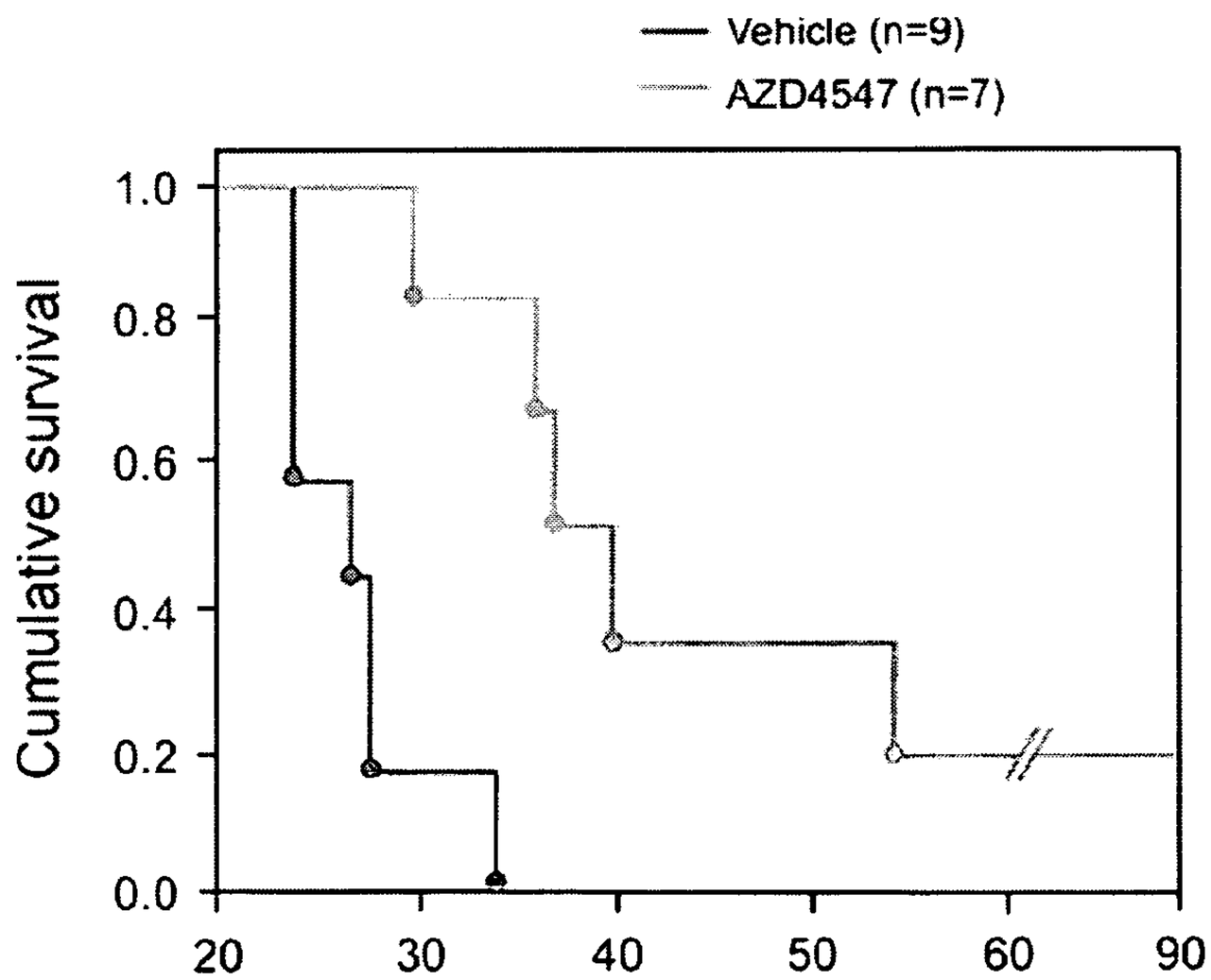

Vehicle (n=9)
AZD4547 (n=7)
Cumulative survival
1.0
0.8
0.6
0.4
0.2
0.0
20
30
40
50
60
90
Days

FIG. 24

MGAPACALALCVAVAIVAGASSESLGTEQRVVGRAAEVPGPEPGQQEQLVFGSGDAVELSCPPPGGGPMGPTVWVKDGTGLVPSERVLVGPQRLQV
LNASHEDSGAYSCRQRLTQRVLCHFSVRVTDAPSSGDDEDGEDEAEDTGVDTGAPYWTRPERMDKKLLAVPAANTVRFRCPAAGNPTPSISWLKNG
REFRGEHRIGGIKLRHQQWSLVMESVVPSDRGNYTCVVENKFGSIRQTYTLDVLERSPHRPILQAGLPANQTAVLGSDVEFHCKVYSDAQPHIQWL
KHVEVNGSKVGPDGTPYVTVLKTAGANTTDKELEVLSLHNVTFEDAGEYTCLAGNSIGFSHHSAWLVVLPAEEELVEADEAGSVYAGILSYGVGFF
LFILVVAAVTLCRLRSPPKKGLGSPTVHKISRFPLKRQVSLESNASMSSNTPLVRIARLSSGEGPTLANVSELELPADPKWELSRARLTLGKPLGE
GCFGQVVMAEAIGIDKDRAAKPVTVAVKMLKDDATDKDLSDLVSEMEMMKMIGKHKNIINLLGACTQGGPLYVLVEYAAKGNLREFLRARRPPGLD
YSFDTCKPPEEQLTFKDLVSCAYQVARGMEYLASQKCIHRDLAARNVLVTEDNVMKIADFGLARDVHNLDYYKKTTNGRLPVKWMAPEALFDRVYT
HQSDVWSFGVLLWEIFTLGGSPYPGIPVEELFKLLKEGHRMDKPANCTHDLYMIMRECWHAAPSQRPTFKQLVEDLDRVLTVTSTDFKESALRKQS
LYLKFDPLLRDSPGRPVPVATETSSMHGANETPSGRPREAKLVEFDFLGALDIPVPGPPPGVPAPGGPPLSTGPIVDLLQYSQKDLDAVVKATQEE
NRELRSRCEELHGKNLELGKIMDRFEEVVYQAMEEVQKQKELSKAEIQKVLKEKDQLTTDLNSMEKSFSDLFKRFEKQKEVIEGYRKNEESLKKCV
EDYLARITQEGQRYQALKAHAEEKLQLANEEIAQVRSKAQAEALALQASLRKEQMRIQSLEKTVEQKTKENEELTRICDDLISKMEKI

FIG.25

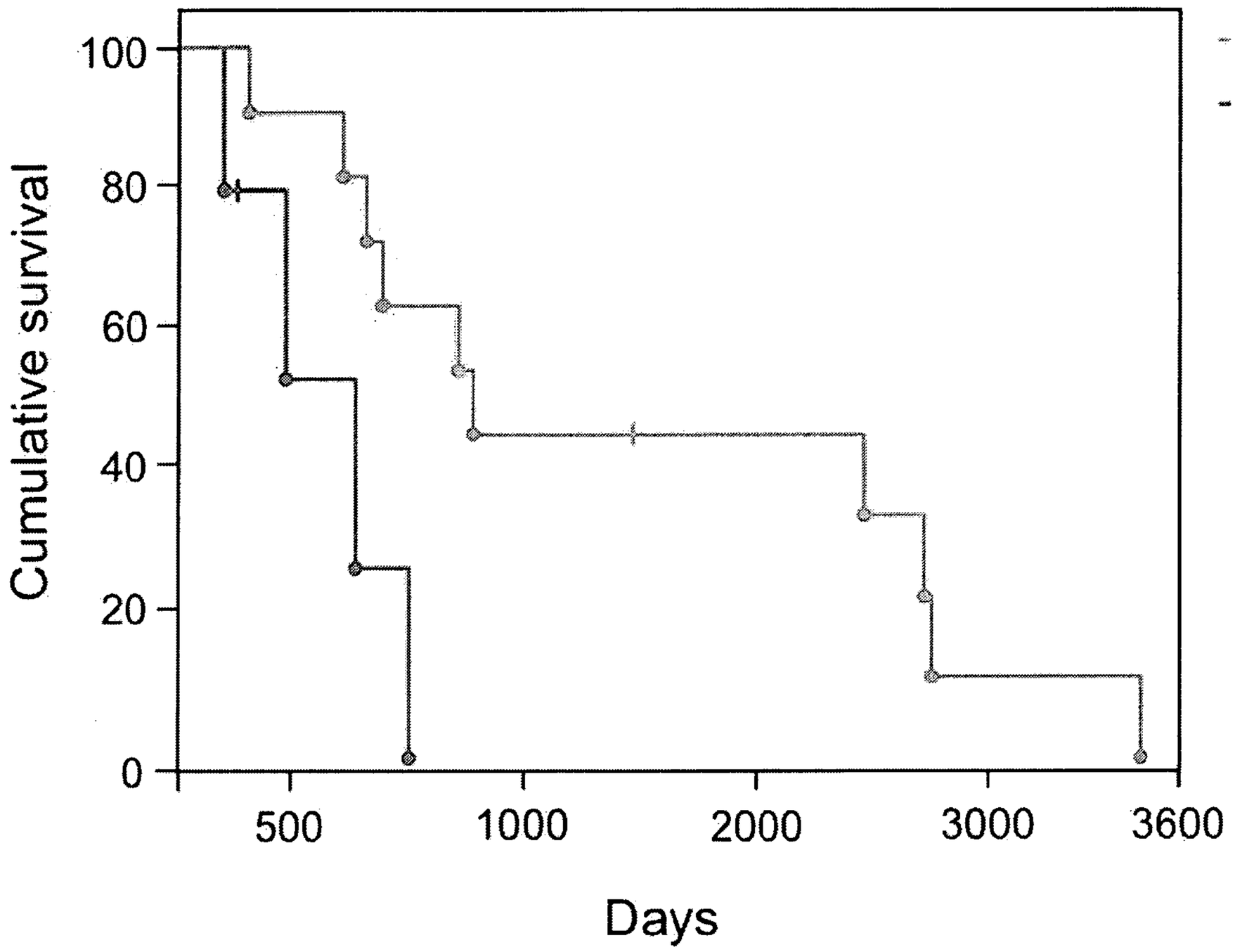

100
80
60
40
20
0
Cumulative survival
500
1000
2000
3000
3600
Days
IDH Mutant (n=11)
FGFR3-TACC3 positive (n=5)

FIG. 26

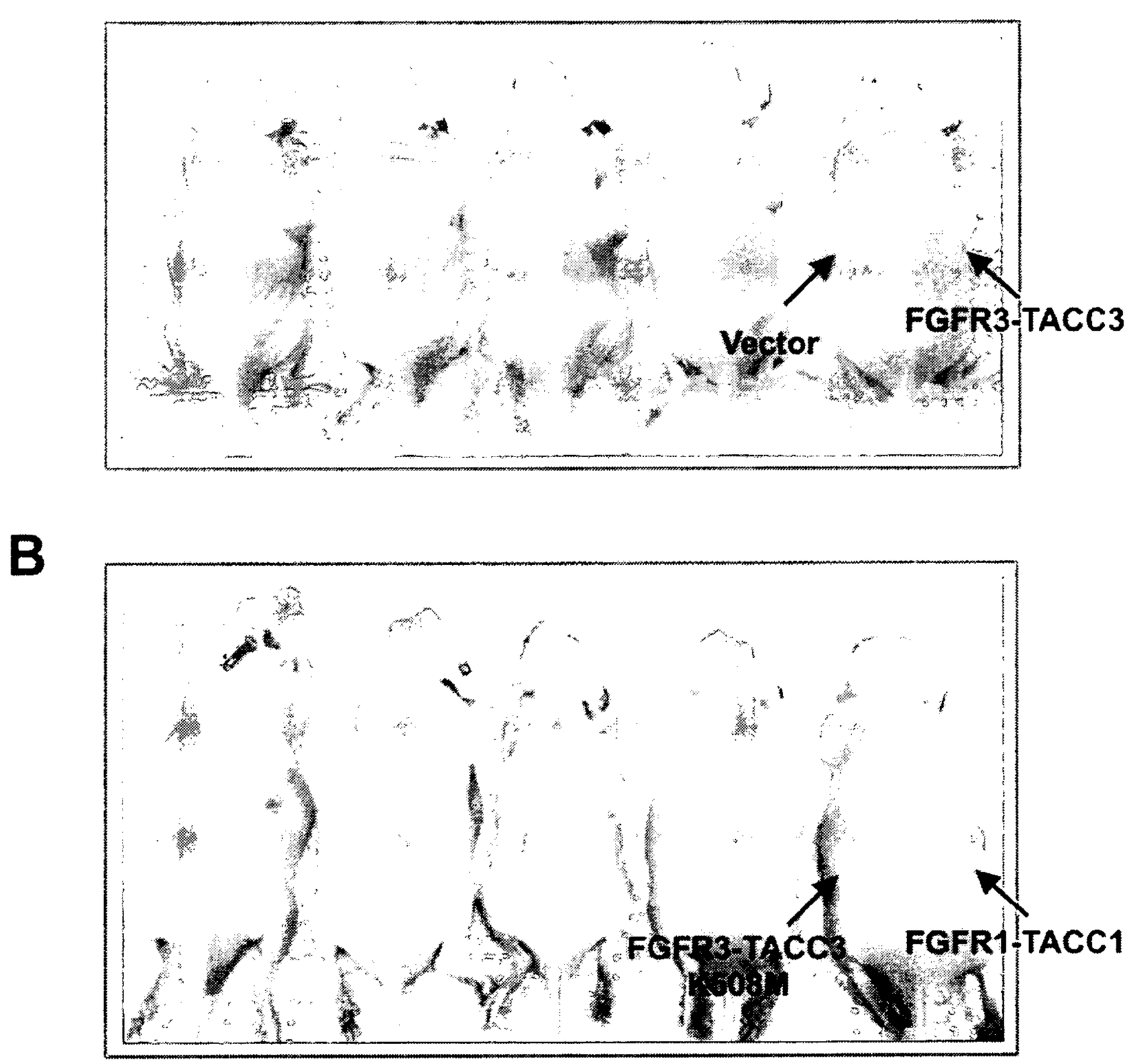


FIG. 27

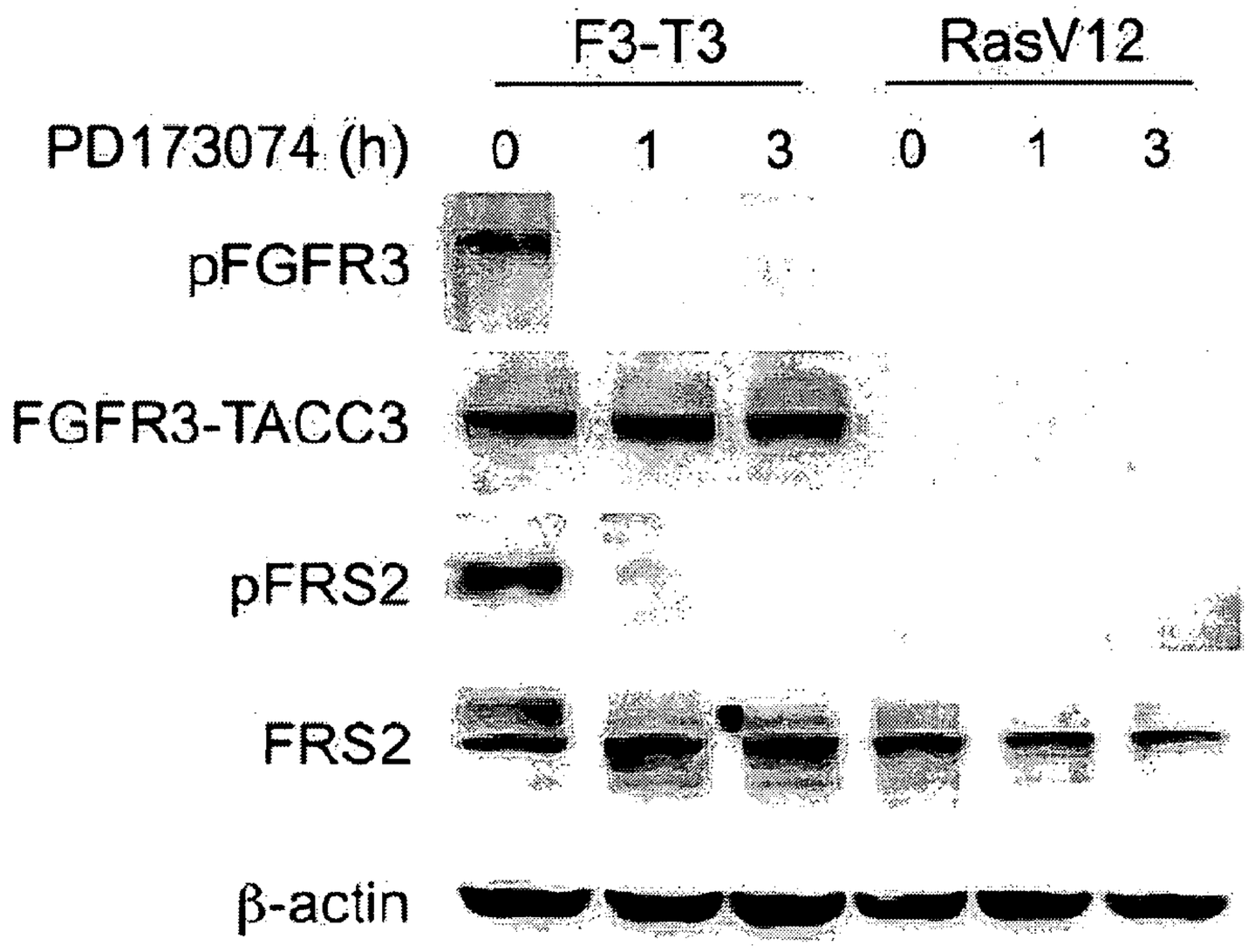

F3-T3
RasV12
PD173074 (h)
0
1
3
0
1
3
pFGFR3
FGFR3-TACC3
pFRS2
FRS2
β-actin

FIG. 28

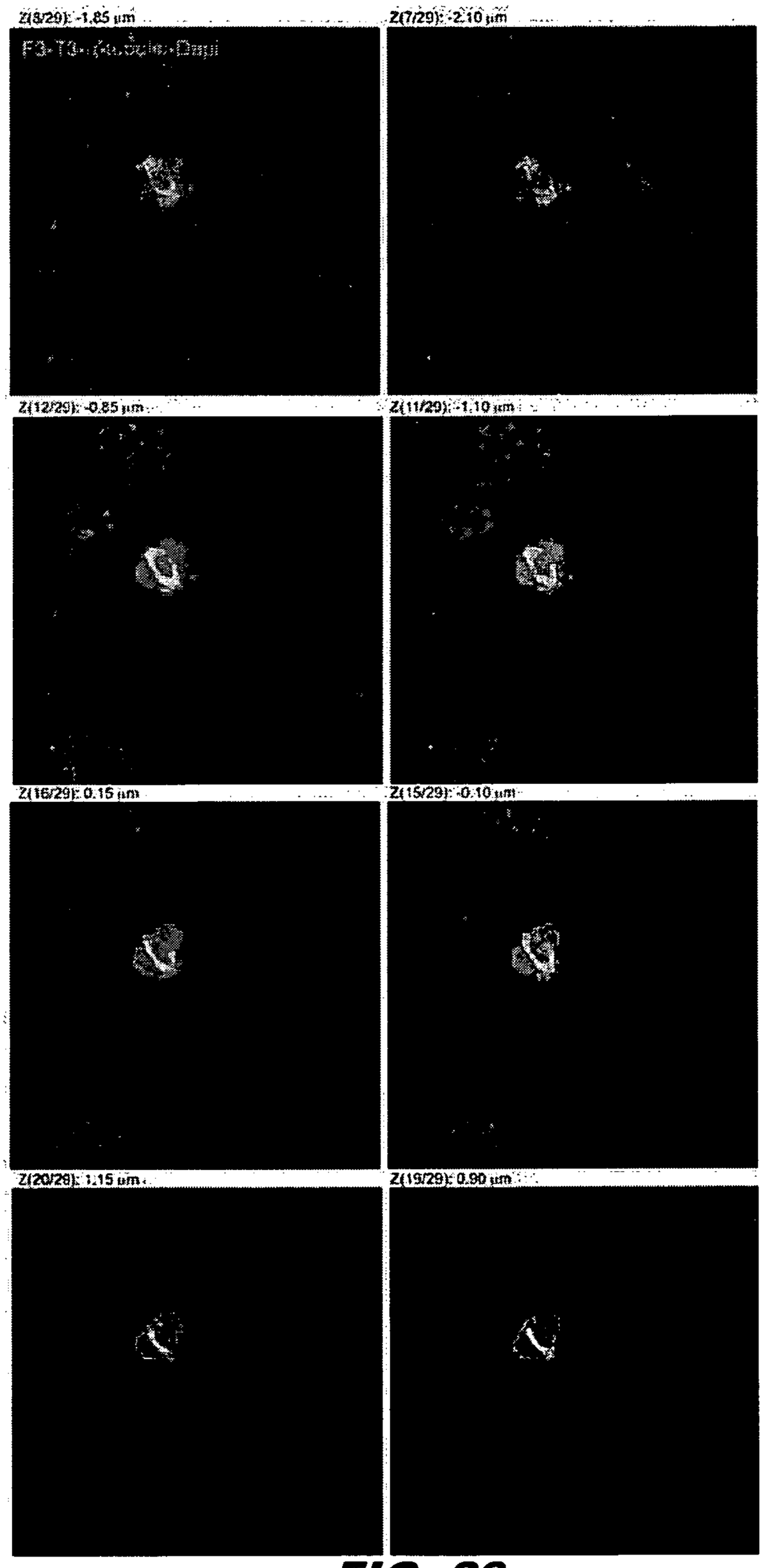

Z(8/29): -1.85 µm
Z(7/29): -2.10 µm
Z(12/29): -0.85 µm
Z(11/29): -1.10 µm
Z(16/29): 0.15 µm
Z(15/29): -0.10 µm
Z(20/29): 1.15 µm
Z(19/29): 0.90 µm

FIG. 29

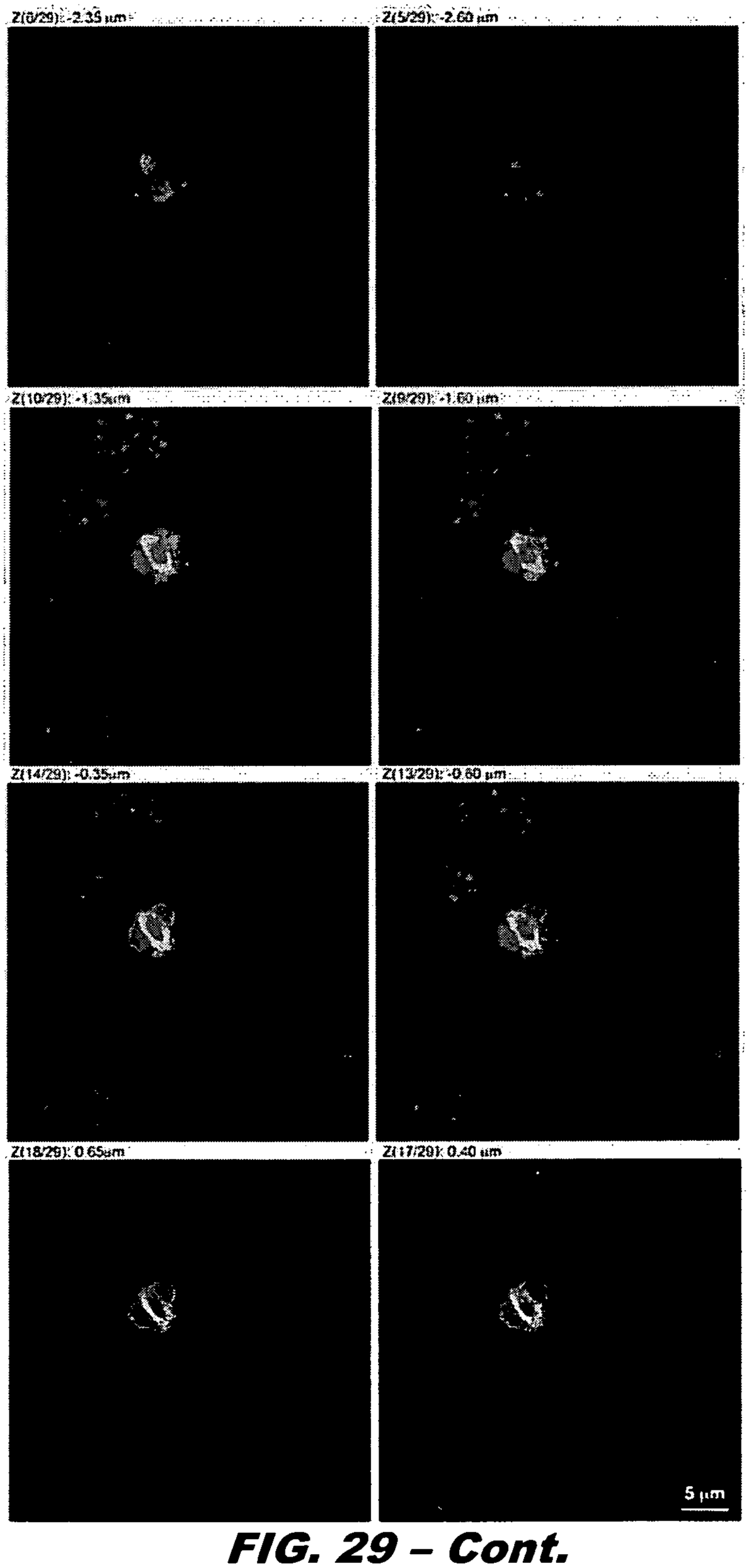

Z(0/29): -2.35 µm
Z(5/29): -2.60 µm
Z(10/29): -1.35µm
Z(9/29): -1.60 µm
Z(14/29): -0.35µm
Z(13/29): -0.60 µm
Z(18/29): 0.65µm
Z(17/29): 0.40 µm
5 µm

FIG. 29 – Cont.

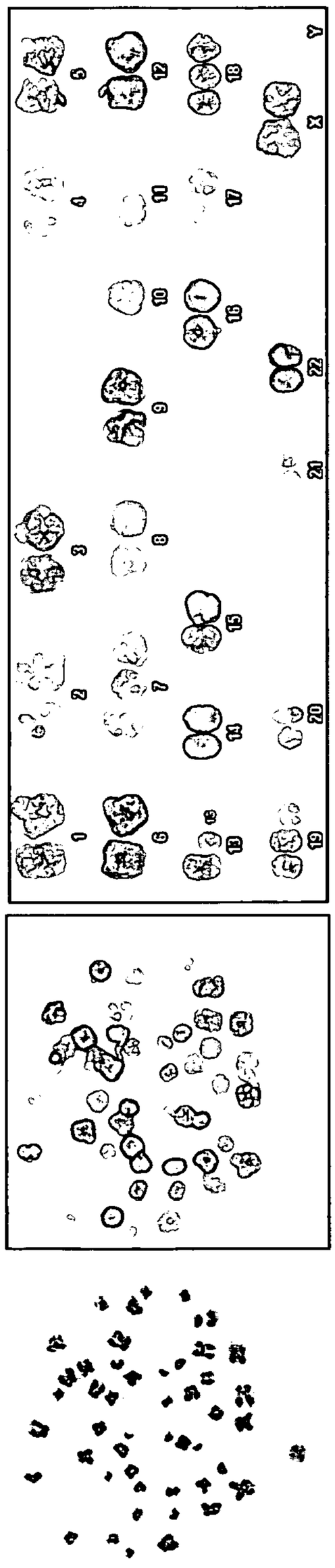


FIG. 30

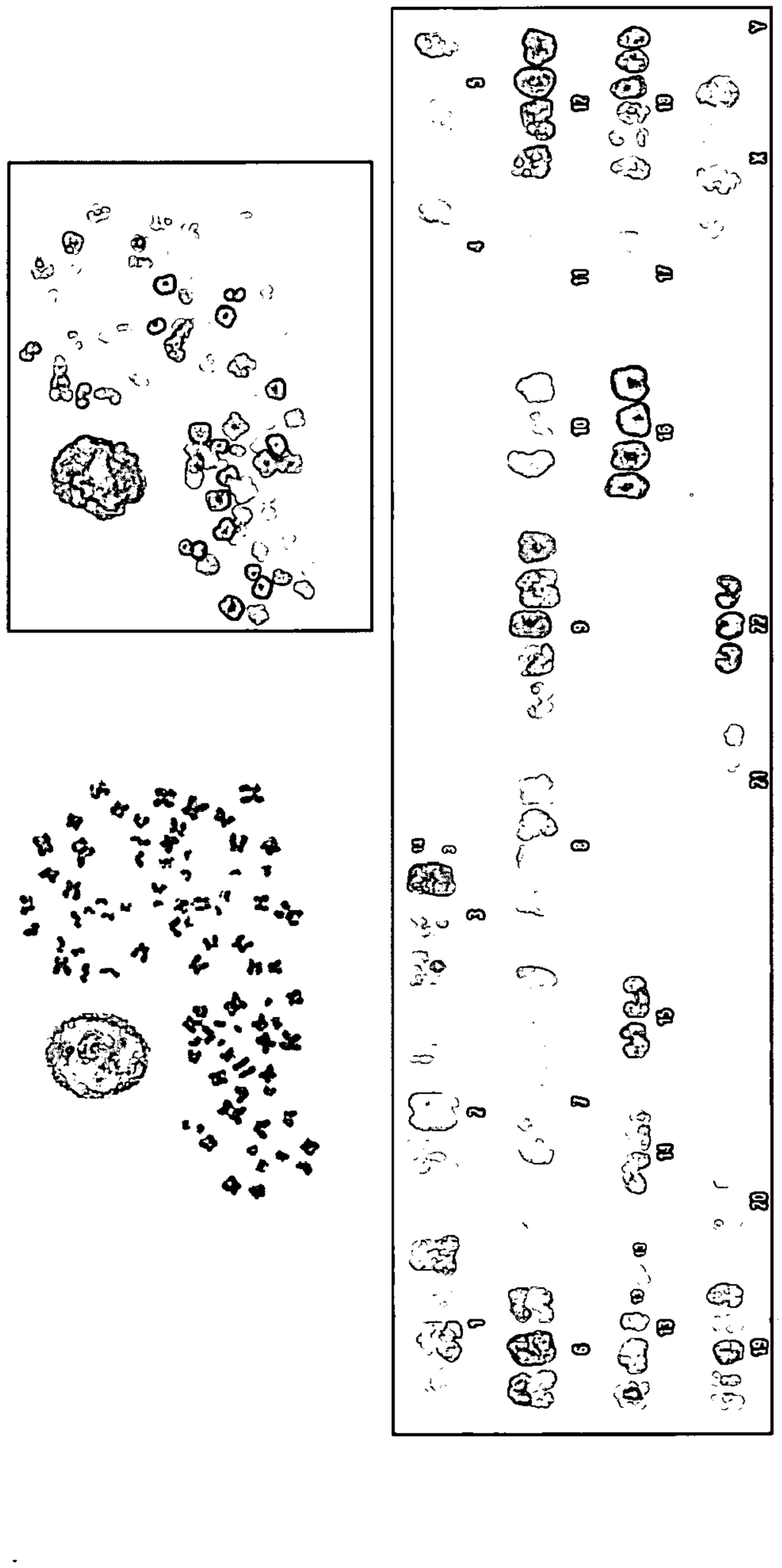


FIG. 30 – CONT.

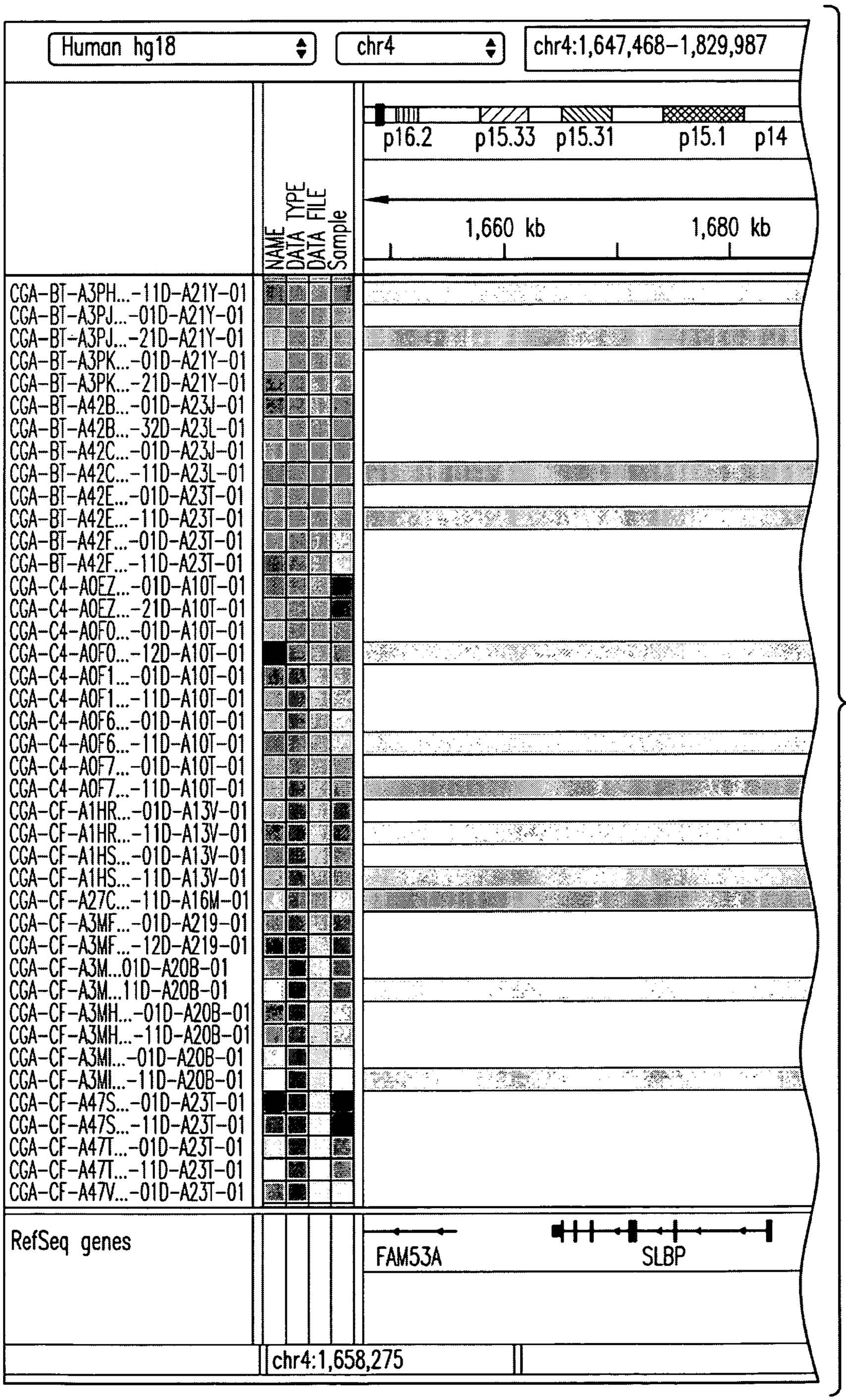

Human hg18
chr4
chr4:1,647,468–1,829,987
p16.2
p15.33
p15.31
p15.1
p14
1,660 kb
1,680 kb
NAME
DATA TYPE
DATA FILE
Sample
CGA-BT-A3PH...-11D-A21Y-01
CGA-BT-A3PJ...-01D-A21Y-01
CGA-BT-A3PJ...-21D-A21Y-01
CGA-BT-A3PK...-01D-A21Y-01
CGA-BT-A3PK...-21D-A21Y-01
CGA-BT-A42B...-01D-A23J-01
CGA-BT-A42B...-32D-A23L-01
CGA-BT-A42C...-01D-A23J-01
CGA-BT-A42C...-11D-A23L-01
CGA-BT-A42E...-01D-A23T-01
CGA-BT-A42E...-11D-A23T-01
CGA-BT-A42F...-01D-A23T-01
CGA-BT-A42F...-11D-A23T-01
CGA-C4-A0EZ...-01D-A10T-01
CGA-C4-A0EZ...-21D-A10T-01
CGA-C4-A0F0...-01D-A10T-01
CGA-C4-A0F0...-12D-A10T-01
CGA-C4-A0F1...-01D-A10T-01
CGA-C4-A0F1...-11D-A10T-01
CGA-C4-A0F6...-01D-A10T-01
CGA-C4-A0F6...-11D-A10T-01
CGA-C4-A0F7...-01D-A10T-01
CGA-C4-A0F7...-11D-A10T-01
CGA-CF-A1HR...-01D-A13V-01
CGA-CF-A1HR...-11D-A13V-01
CGA-CF-A1HS...-01D-A13V-01
CGA-CF-A1HS...-11D-A13V-01
CGA-CF-A27C...-11D-A16M-01
CGA-CF-A3MF...-01D-A219-01
CGA-CF-A3MF...-12D-A219-01
CGA-CF-A3M...01D-A20B-01
CGA-CF-A3M...11D-A20B-01
CGA-CF-A3MH...-01D-A20B-01
CGA-CF-A3MH...-11D-A20B-01
CGA-CF-A3MI...-01D-A20B-01
CGA-CF-A3MI...-11D-A20B-01
CGA-CF-A47S...-01D-A23T-01
CGA-CF-A47S...-11D-A23T-01
CGA-CF-A47T...-01D-A23T-01
CGA-CF-A47T...-11D-A23T-01
CGA-CF-A47V...-01D-A23T-01
Cont. on FIG.31-2
RefSeq genes
FAM53A
SLBP
chr4:1,658,275

FIG.31-1

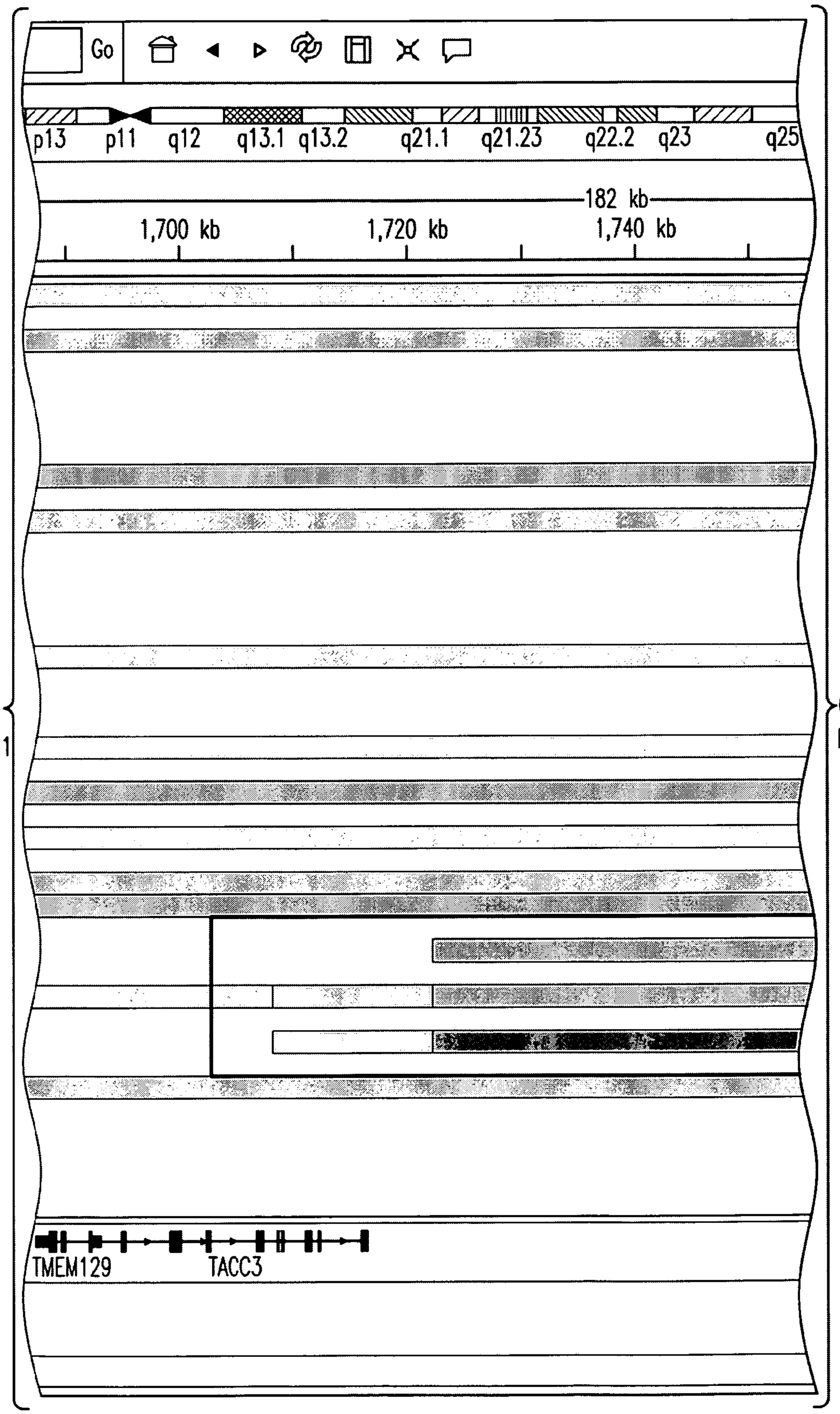

Go
p13
p11
q12
q13.1
q13.2
q21.1
q21.23
q22.2
q23
q25
182 kb
1,700 kb
1,720 kb
1,740 kb
Cont. From FIG.31-1
Cont. on FIG.31-3
TMEM129
TACC3

FIG.31-2

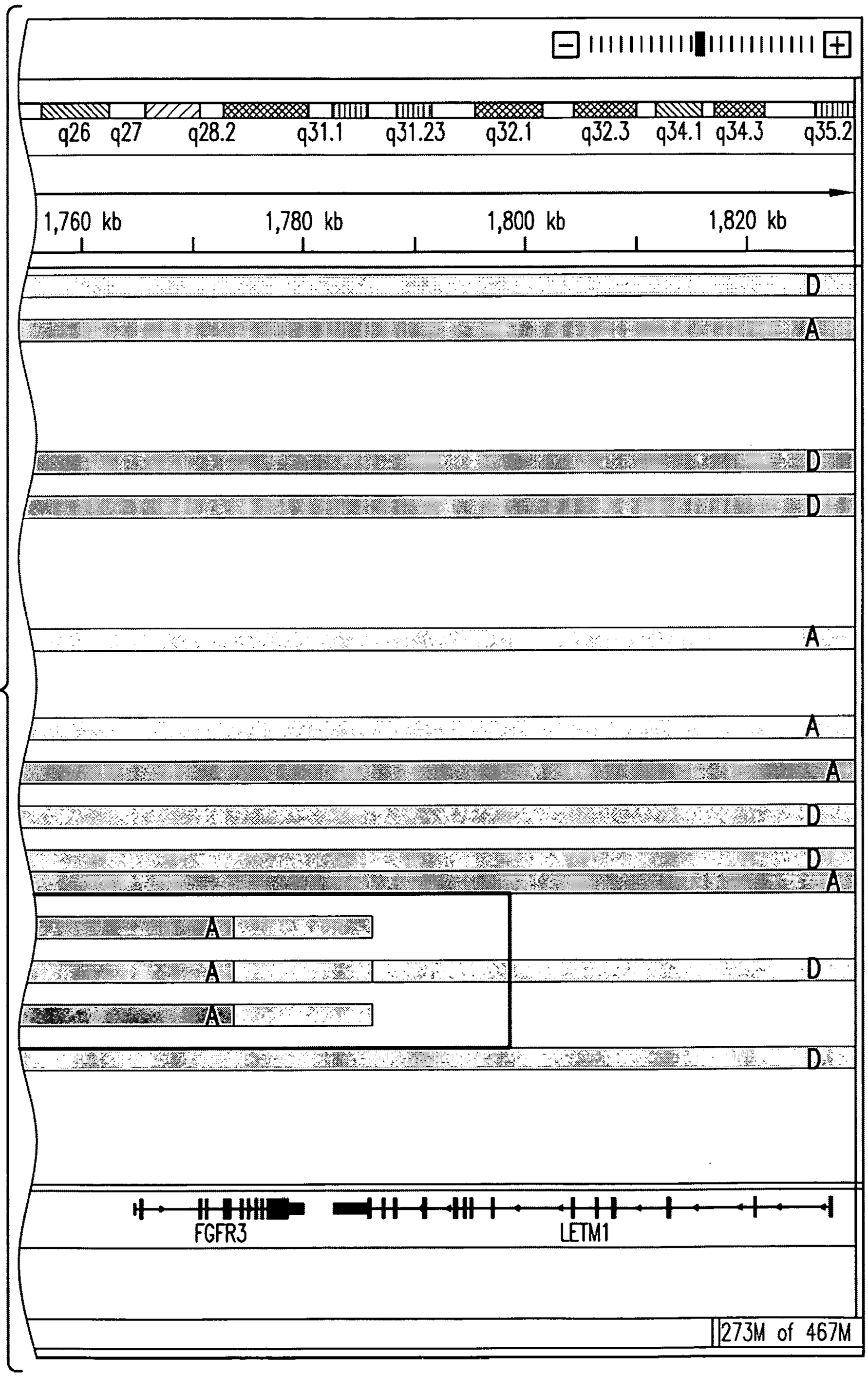

q26
q27
q28.2
q31.1
q31.23
q32.1
q32.3
q34.1
q34.3
q35.2
1,760 kb
1,780 kb
1,800 kb
1,820 kb
D
A
D
D
Cont.
From
FIG.31-2
A
A
A
D
D
A
A
A
D
D
FGFR3
LETM1
273M of 467M

FIG.31-3

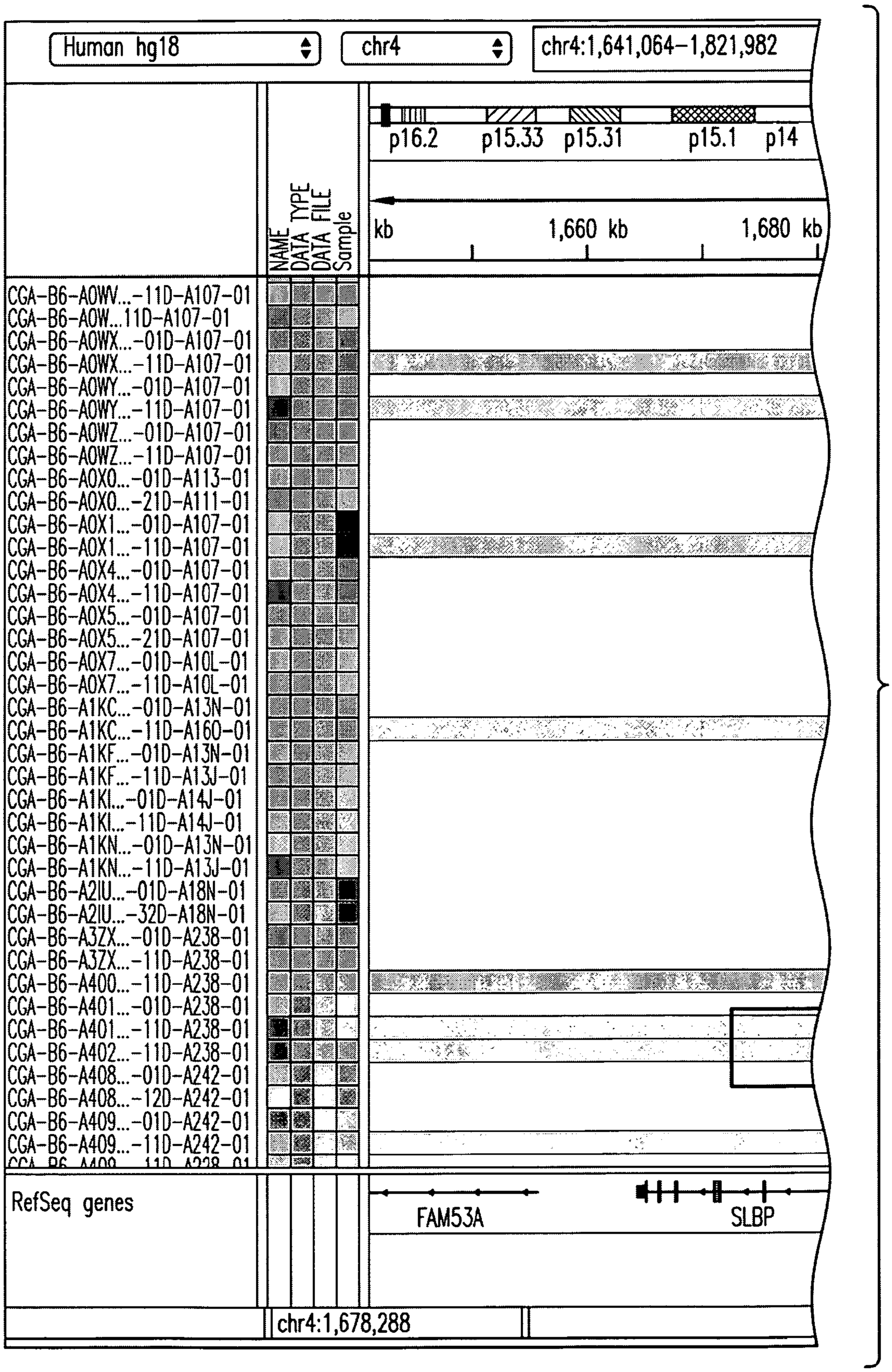

Human hg18
chr4
chr4:1,641,064–1,821,982
p16.2
p15.33
p15.31
p15.1
p14
kb
1,660 kb
1,680 kb
NAME
DATA TYPE
DATA FILE
Sample
CGA-B6-A0WV...-11D-A107-01
CGA-B6-A0W...11D-A107-01
CGA-B6-A0WX...-01D-A107-01
CGA-B6-A0WX...-11D-A107-01
CGA-B6-A0WY...-01D-A107-01
CGA-B6-A0WY...-11D-A107-01
CGA-B6-A0WZ...-01D-A107-01
CGA-B6-A0WZ...-11D-A107-01
CGA-B6-A0X0...-01D-A113-01
CGA-B6-A0X0...-21D-A111-01
CGA-B6-A0X1...-01D-A107-01
CGA-B6-A0X1...-11D-A107-01
CGA-B6-A0X4...-01D-A107-01
CGA-B6-A0X4...-11D-A107-01
CGA-B6-A0X5...-01D-A107-01
CGA-B6-A0X5...-21D-A107-01
CGA-B6-A0X7...-01D-A10L-01
CGA-B6-A0X7...-11D-A10L-01
CGA-B6-A1KC...-01D-A13N-01
CGA-B6-A1KC...-11D-A160-01
CGA-B6-A1KF...-01D-A13N-01
CGA-B6-A1KF...-11D-A13J-01
CGA-B6-A1KI...-01D-A14J-01
CGA-B6-A1KI...-11D-A14J-01
CGA-B6-A1KN...-01D-A13N-01
CGA-B6-A1KN...-11D-A13J-01
CGA-B6-A2IU...-01D-A18N-01
CGA-B6-A2IU...-32D-A18N-01
CGA-B6-A3ZX...-01D-A238-01
CGA-B6-A3ZX...-11D-A238-01
CGA-B6-A400...-11D-A238-01
CGA-B6-A401...-01D-A238-01
CGA-B6-A401...-11D-A238-01
CGA-B6-A402...-11D-A238-01
CGA-B6-A408...-01D-A242-01
CGA-B6-A408...-12D-A242-01
CGA-B6-A409...-01D-A242-01
CGA-B6-A409...-11D-A242-01
Cont. on FIG.32-2
RefSeq genes
FAM53A
SLBP
chr4:1,678,288

FIG.32-1

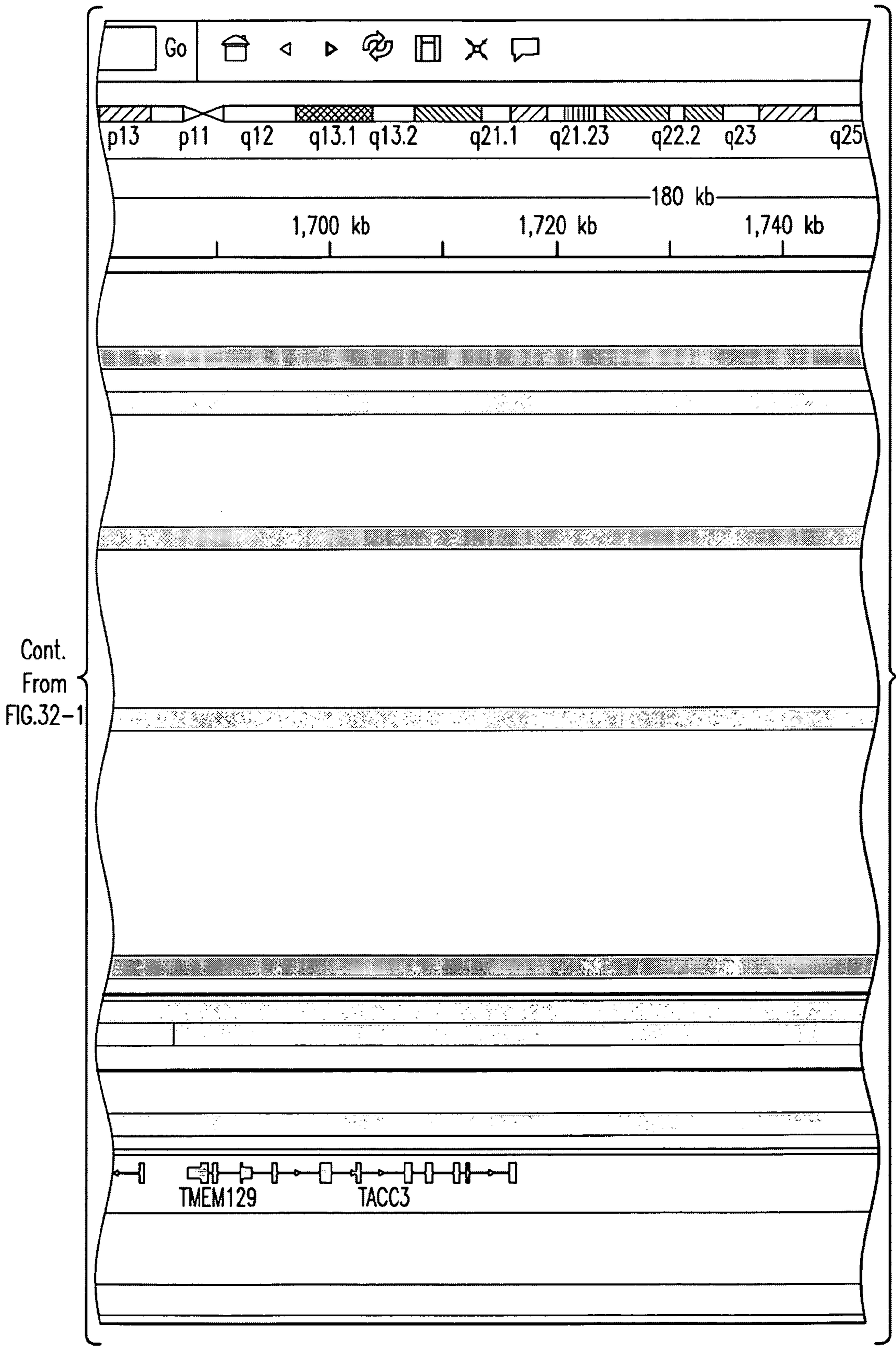

Go
p13
p11
q12
q13.1
q13.2
q21.1
q21.23
q22.2
q23
q25
180 kb
1,700 kb
1,720 kb
1,740 kb
Cont. From FIG.32-1
Cont. on FIG.32-3
TMEM129
TACC3

FIG.32-2

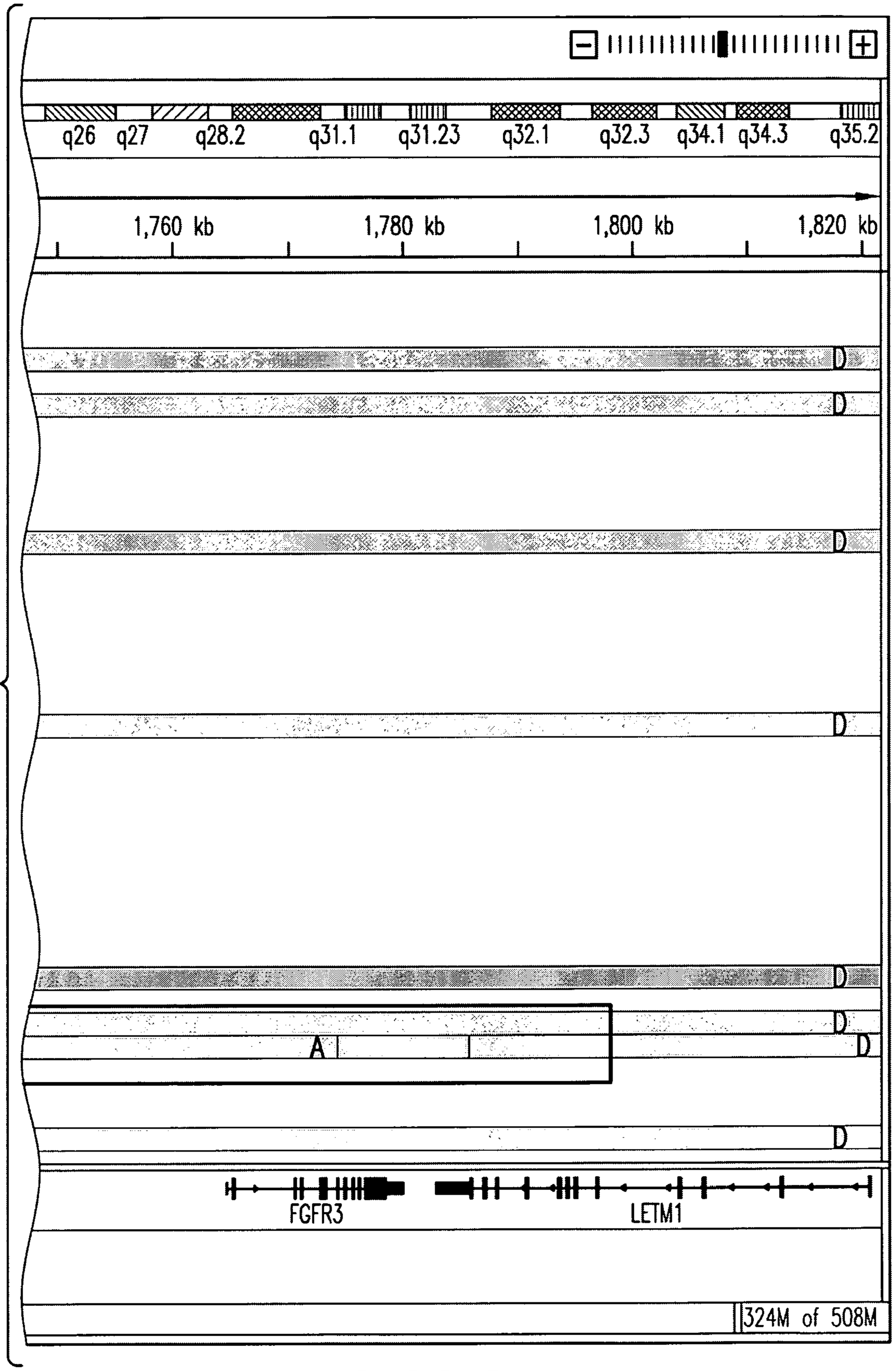

q26
q27
q28.2
q31.1
q31.23
q32.1
q32.3
q34.1
q34.3
q35.2
1,760 kb
1,780 kb
1,800 kb
1,820 kb
Cont.
From
FIG.32-2
D
A
FGFR3
LETM1
324M of 508M

FIG.32-3

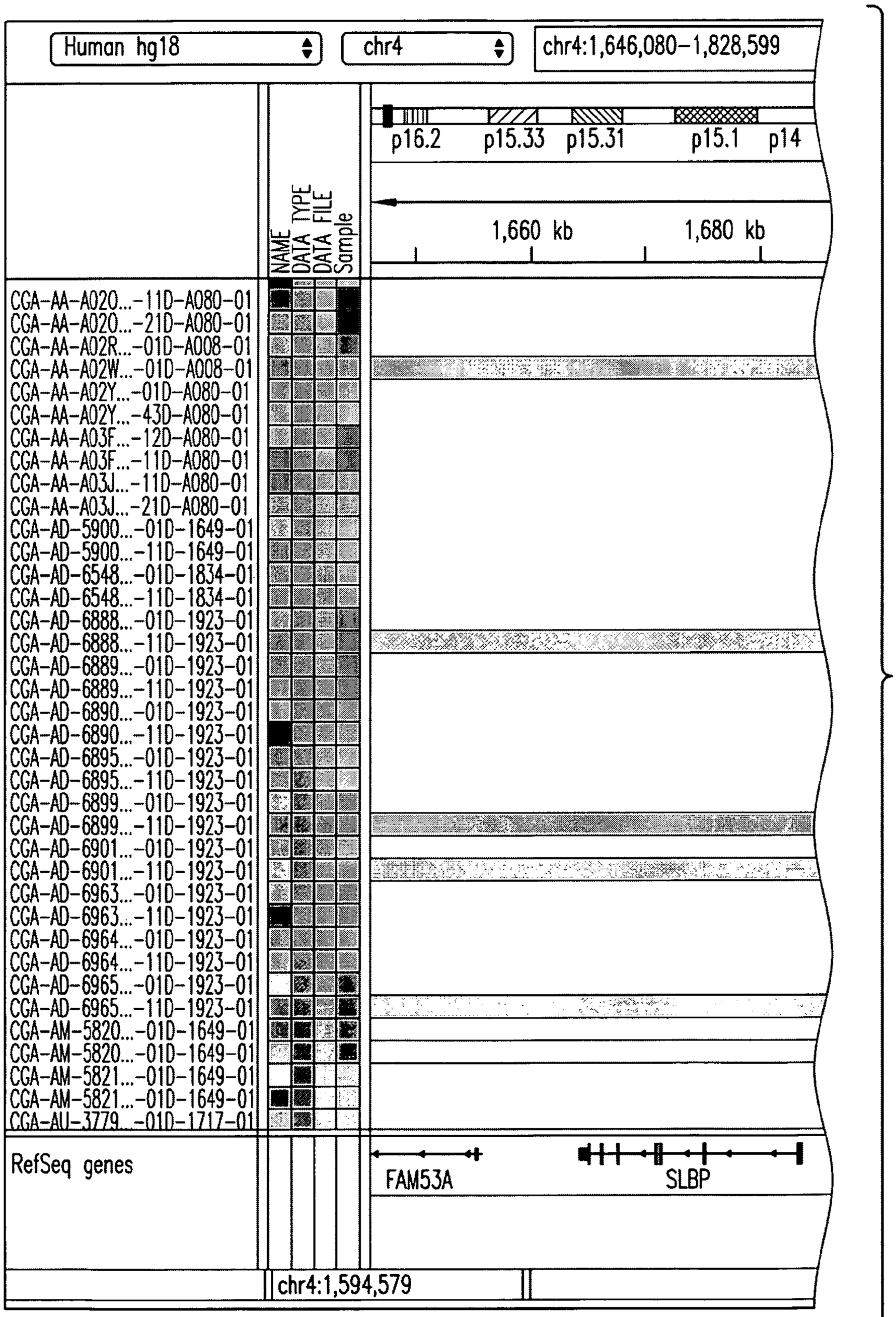

Human hg18
chr4
chr4:1,646,080–1,828,599
p16.2
p15.33
p15.31
p15.1
p14
1,660 kb
1,680 kb
NAME
DATA TYPE
DATA FILE
Sample
CGA-AA-A020...-11D-A080-01
CGA-AA-A020...-21D-A080-01
CGA-AA-A02R...-01D-A008-01
CGA-AA-A02W...-01D-A008-01
CGA-AA-A02Y...-01D-A080-01
CGA-AA-A02Y...-43D-A080-01
CGA-AA-A03F...-12D-A080-01
CGA-AA-A03F...-11D-A080-01
CGA-AA-A03J...-11D-A080-01
CGA-AA-A03J...-21D-A080-01
CGA-AD-5900...-01D-1649-01
CGA-AD-5900...-11D-1649-01
CGA-AD-6548...-01D-1834-01
CGA-AD-6548...-11D-1834-01
CGA-AD-6888...-01D-1923-01
CGA-AD-6888...-11D-1923-01
CGA-AD-6889...-01D-1923-01
CGA-AD-6889...-11D-1923-01
CGA-AD-6890...-01D-1923-01
CGA-AD-6890...-11D-1923-01
CGA-AD-6895...-01D-1923-01
CGA-AD-6895...-11D-1923-01
CGA-AD-6899...-01D-1923-01
CGA-AD-6899...-11D-1923-01
CGA-AD-6901...-01D-1923-01
CGA-AD-6901...-11D-1923-01
CGA-AD-6963...-01D-1923-01
CGA-AD-6963...-11D-1923-01
CGA-AD-6964...-01D-1923-01
CGA-AD-6964...-11D-1923-01
CGA-AD-6965...-01D-1923-01
CGA-AD-6965...-11D-1923-01
CGA-AM-5820...-01D-1649-01
CGA-AM-5820...-01D-1649-01
CGA-AM-5821...-01D-1649-01
CGA-AM-5821...-01D-1649-01
CGA-AU-3779...-01D-1717-01
RefSeq genes
FAM53A
SLBP
chr4:1,594,579
Cont. on FIG.33-2

FIG.33-1

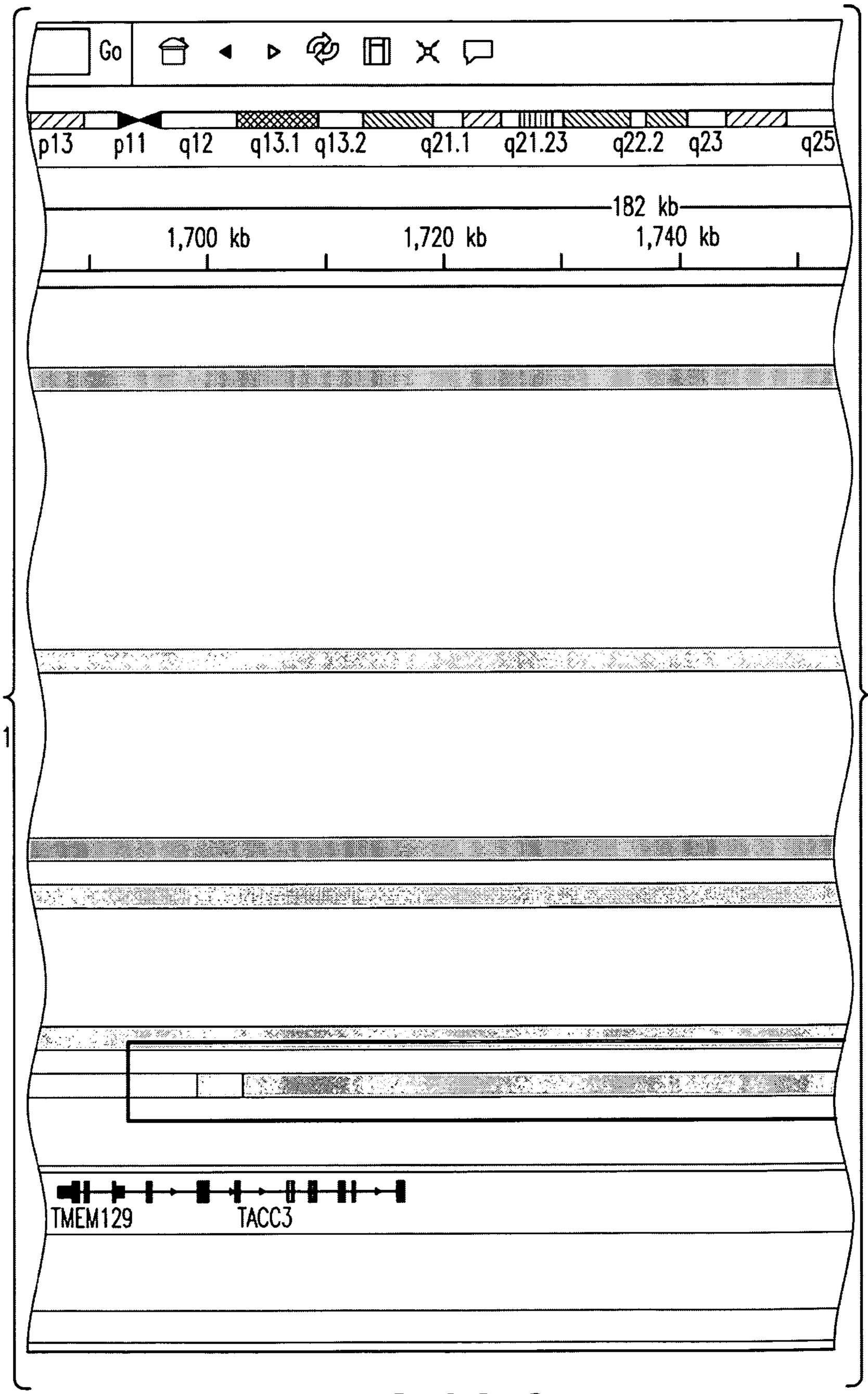

Go
p13
p11
q12
q13.1
q13.2
q21.1
q21.23
q22.2
q23
q25
182 kb
1,700 kb
1,720 kb
1,740 kb
Cont. From FIG.33-1
Cont. on FIG.33-3
TMEM129
TACC3

FIG.33-2

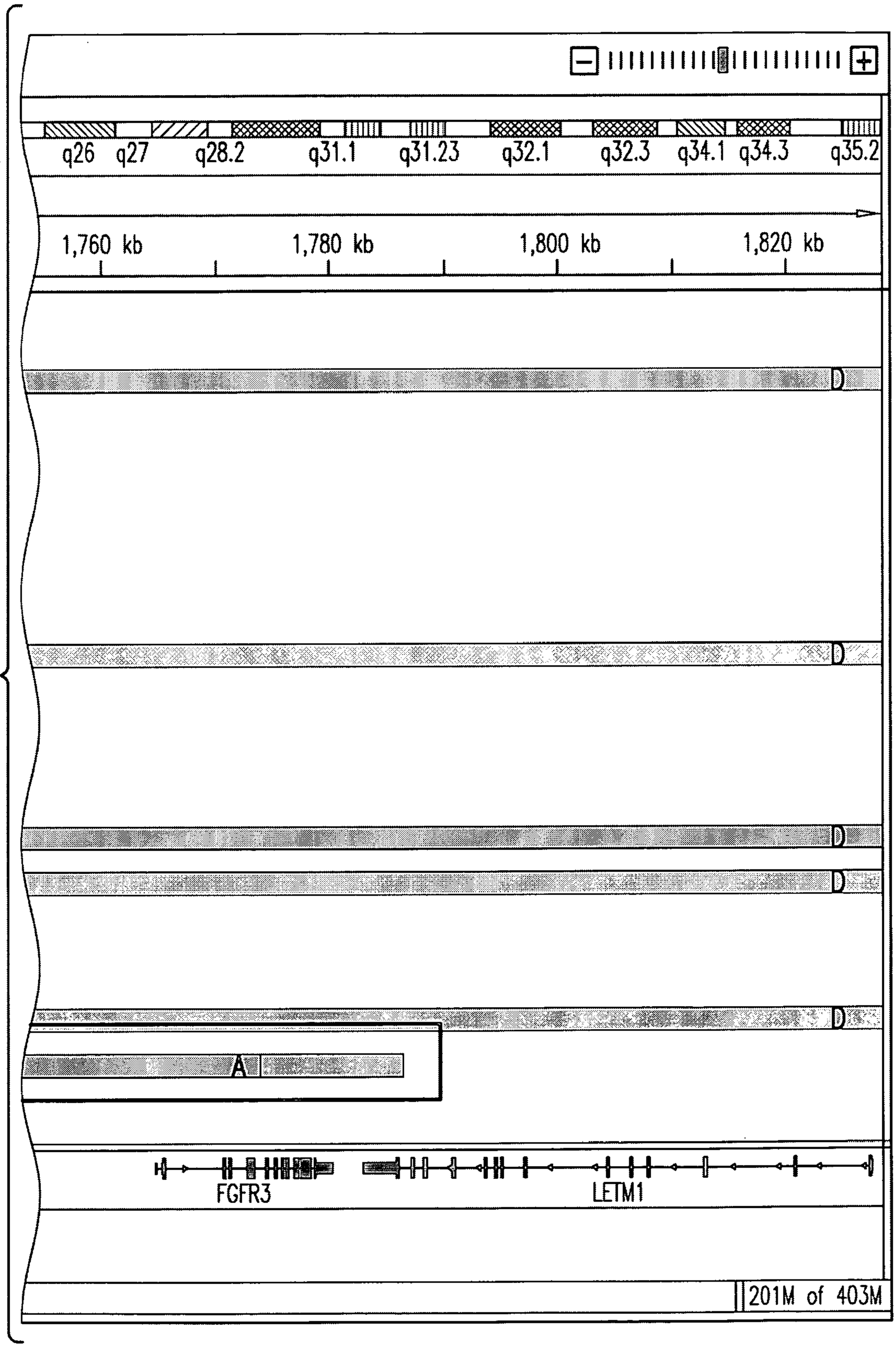

q26
q27
q28.2
q31.1
q31.23
q32.1
q32.3
q34.1
q34.3
q35.2
1,760 kb
1,780 kb
1,800 kb
1,820 kb
Cont.
From
FIG.33-2
A
FGFR3
LETM1
201M of 403M

FIG.33-3

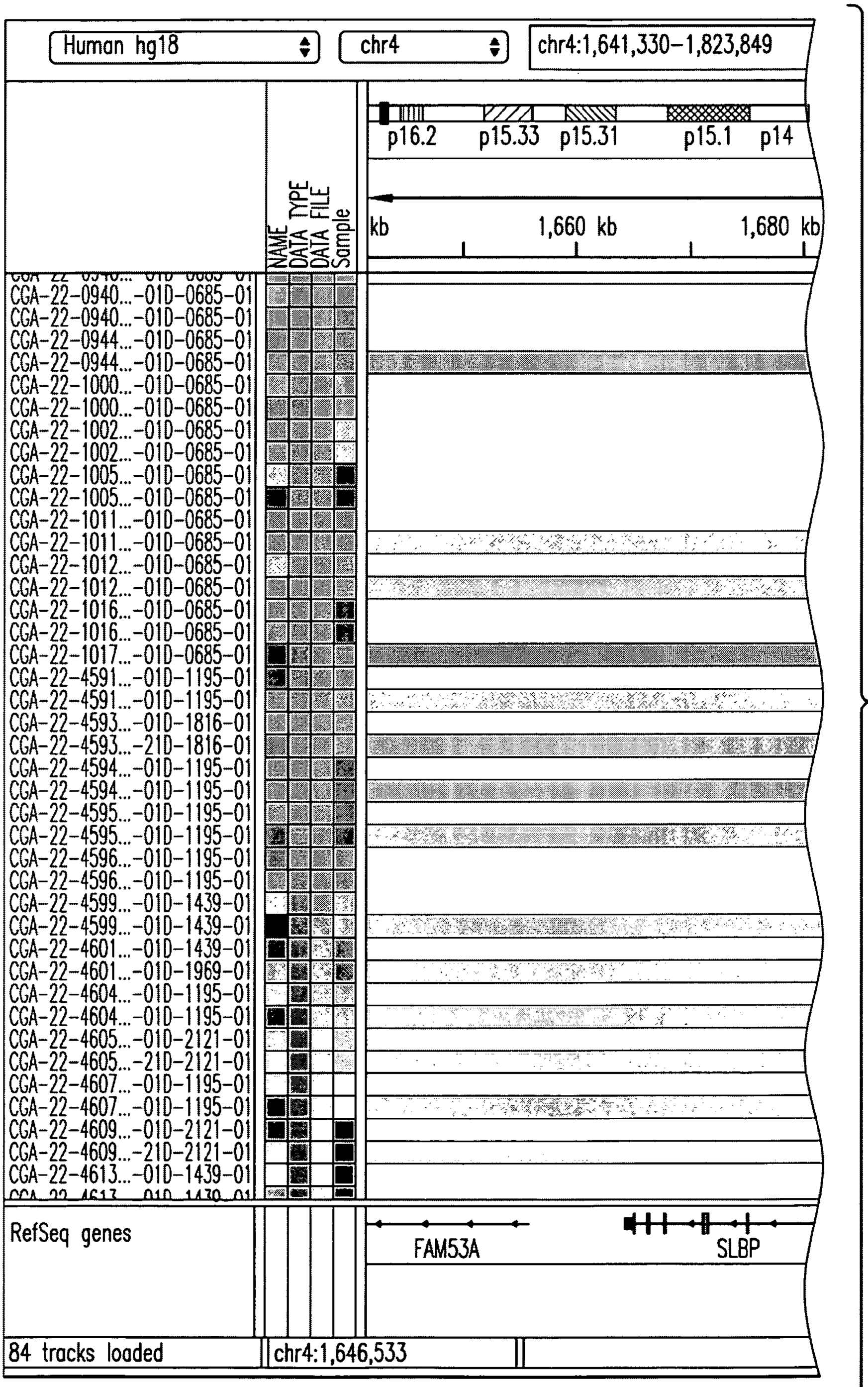

Human hg18
chr4
chr4:1,641,330–1,823,849
p16.2
p15.33
p15.31
p15.1
p14
kb
1,660 kb
1,680 kb
NAME
DATA TYPE
DATA FILE
Sample
CGA-22-0940...-01D-0685-01
CGA-22-0940...-01D-0685-01
CGA-22-0944...-01D-0685-01
CGA-22-0944...-01D-0685-01
CGA-22-1000...-01D-0685-01
CGA-22-1000...-01D-0685-01
CGA-22-1002...-01D-0685-01
CGA-22-1002...-01D-0685-01
CGA-22-1005...-01D-0685-01
CGA-22-1005...-01D-0685-01
CGA-22-1011...-01D-0685-01
CGA-22-1011...-01D-0685-01
CGA-22-1012...-01D-0685-01
CGA-22-1012...-01D-0685-01
CGA-22-1016...-01D-0685-01
CGA-22-1016...-01D-0685-01
CGA-22-1017...-01D-0685-01
CGA-22-4591...-01D-1195-01
CGA-22-4591...-01D-1195-01
CGA-22-4593...-01D-1816-01
CGA-22-4593...-21D-1816-01
CGA-22-4594...-01D-1195-01
CGA-22-4594...-01D-1195-01
CGA-22-4595...-01D-1195-01
CGA-22-4595...-01D-1195-01
CGA-22-4596...-01D-1195-01
CGA-22-4596...-01D-1195-01
CGA-22-4599...-01D-1439-01
CGA-22-4599...-01D-1439-01
CGA-22-4601...-01D-1439-01
CGA-22-4601...-01D-1969-01
CGA-22-4604...-01D-1195-01
CGA-22-4604...-01D-1195-01
CGA-22-4605...-01D-2121-01
CGA-22-4605...-21D-2121-01
CGA-22-4607...-01D-1195-01
CGA-22-4607...-01D-1195-01
CGA-22-4609...-01D-2121-01
CGA-22-4609...-21D-2121-01
CGA-22-4613...-01D-1439-01
Cont. on FIG.34-2
RefSeq genes
FAM53A
SLBP
84 tracks loaded
chr4:1,646,533

FIG.34-1

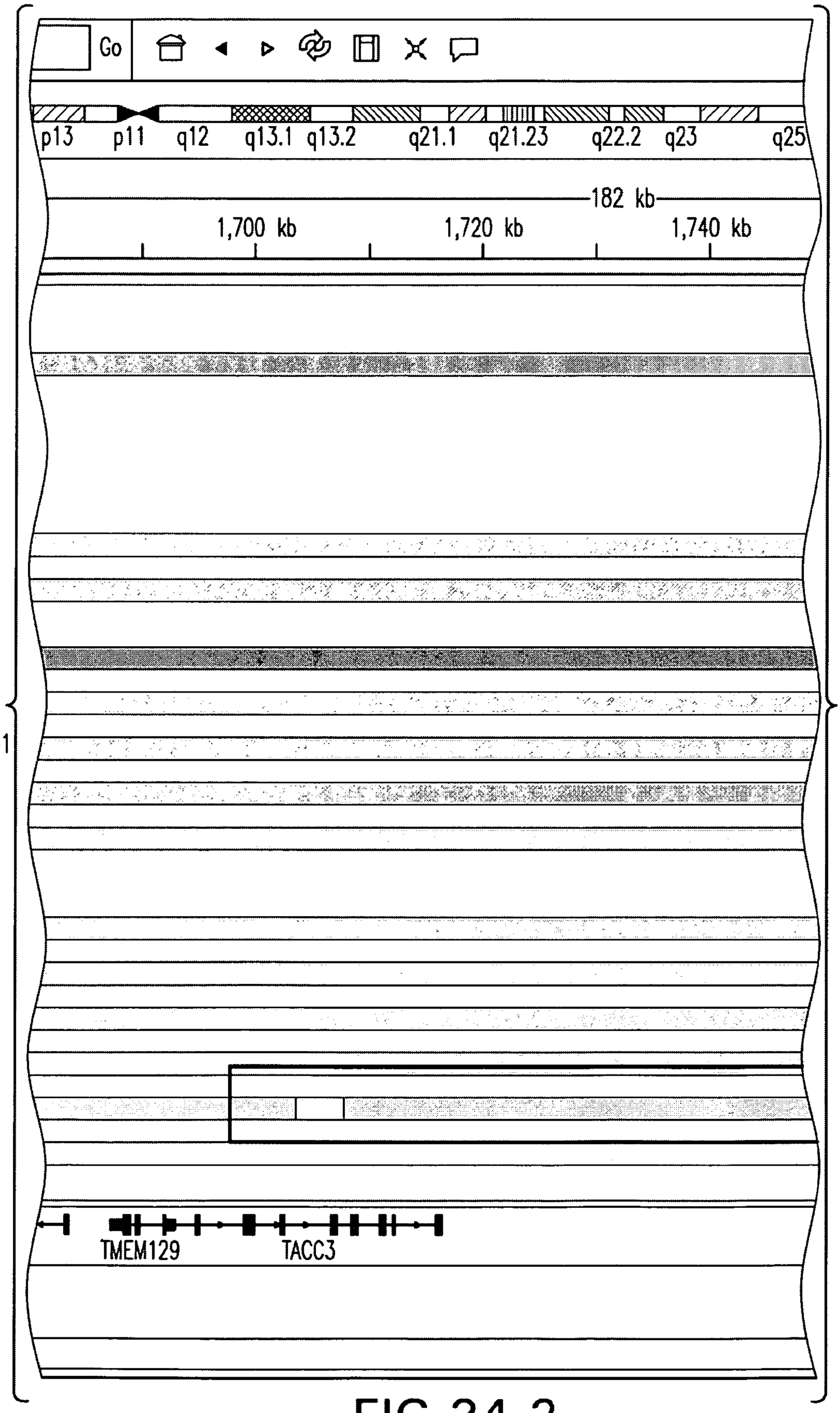

Go
p13
p11
q12
q13.1
q13.2
q21.1
q21.23
q22.2
q23
q25
182 kb
1,700 kb
1,720 kb
1,740 kb
Cont. From FIG.34-1
Cont. on FIG.34-3
TMEM129
TACC3

FIG.34-2

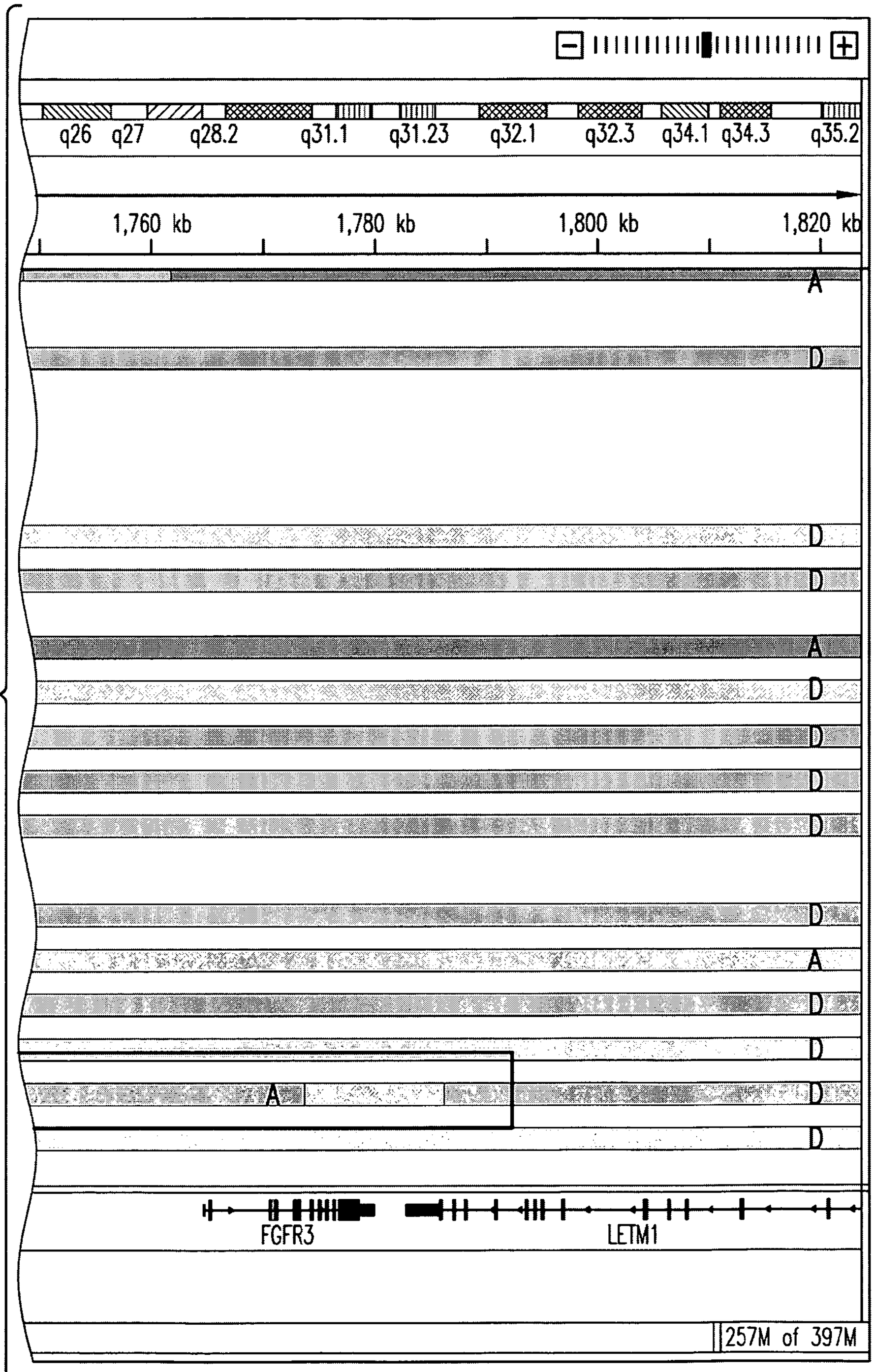

q26
q27
q28.2
q31.1
q31.23
q32.1
q32.3
q34.1
q34.3
q35.2
1,760 kb
1,780 kb
1,800 kb
1,820 kb
A
D
Cont.
From
FIG.34-2
FGFR3
LETM1
257M of 397M

FIG.34-3

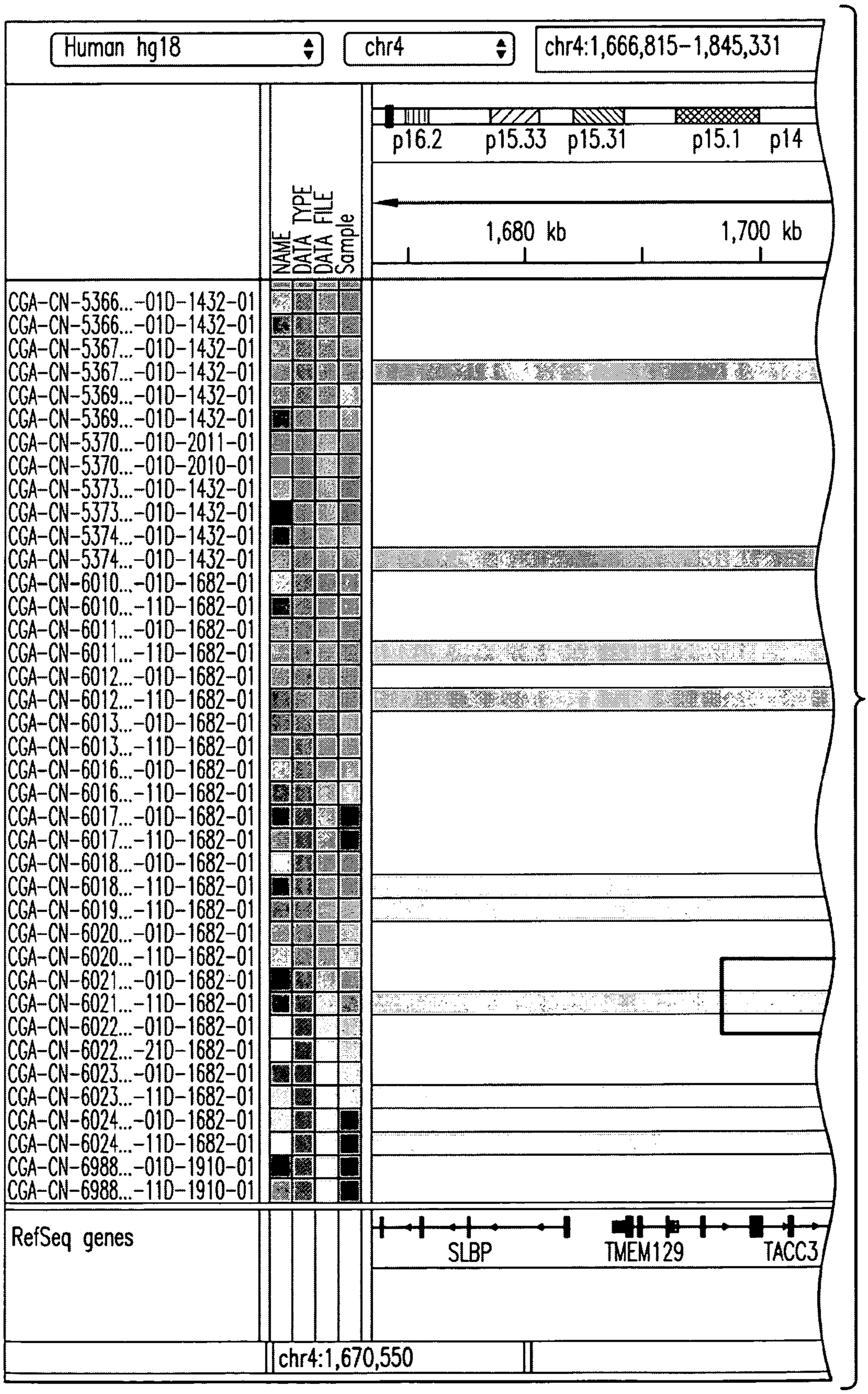

Human hg18
chr4
chr4:1,666,815-1,845,331
p16.2
p15.33
p15.31
p15.1
p14
1,680 kb
1,700 kb
NAME
DATA TYPE
DATA FILE
Sample
CGA-CN-5366...-01D-1432-01
CGA-CN-5366...-01D-1432-01
CGA-CN-5367...-01D-1432-01
CGA-CN-5367...-01D-1432-01
CGA-CN-5369...-01D-1432-01
CGA-CN-5369...-01D-1432-01
CGA-CN-5370...-01D-2011-01
CGA-CN-5370...-01D-2010-01
CGA-CN-5373...-01D-1432-01
CGA-CN-5373...-01D-1432-01
CGA-CN-5374...-01D-1432-01
CGA-CN-5374...-01D-1432-01
CGA-CN-6010...-01D-1682-01
CGA-CN-6010...-11D-1682-01
CGA-CN-6011...-01D-1682-01
CGA-CN-6011...-11D-1682-01
CGA-CN-6012...-01D-1682-01
CGA-CN-6012...-11D-1682-01
CGA-CN-6013...-01D-1682-01
CGA-CN-6013...-11D-1682-01
CGA-CN-6016...-01D-1682-01
CGA-CN-6016...-11D-1682-01
CGA-CN-6017...-01D-1682-01
CGA-CN-6017...-11D-1682-01
CGA-CN-6018...-01D-1682-01
CGA-CN-6018...-11D-1682-01
CGA-CN-6019...-11D-1682-01
CGA-CN-6020...-01D-1682-01
CGA-CN-6020...-11D-1682-01
CGA-CN-6021...-01D-1682-01
CGA-CN-6021...-11D-1682-01
CGA-CN-6022...-01D-1682-01
CGA-CN-6022...-21D-1682-01
CGA-CN-6023...-01D-1682-01
CGA-CN-6023...-11D-1682-01
CGA-CN-6024...-01D-1682-01
CGA-CN-6024...-11D-1682-01
CGA-CN-6988...-01D-1910-01
CGA-CN-6988...-11D-1910-01
Cont. on FIG.35-2
RefSeq genes
SLBP
TMEM129
TACC3
chr4:1,670,550

FIG.35-1

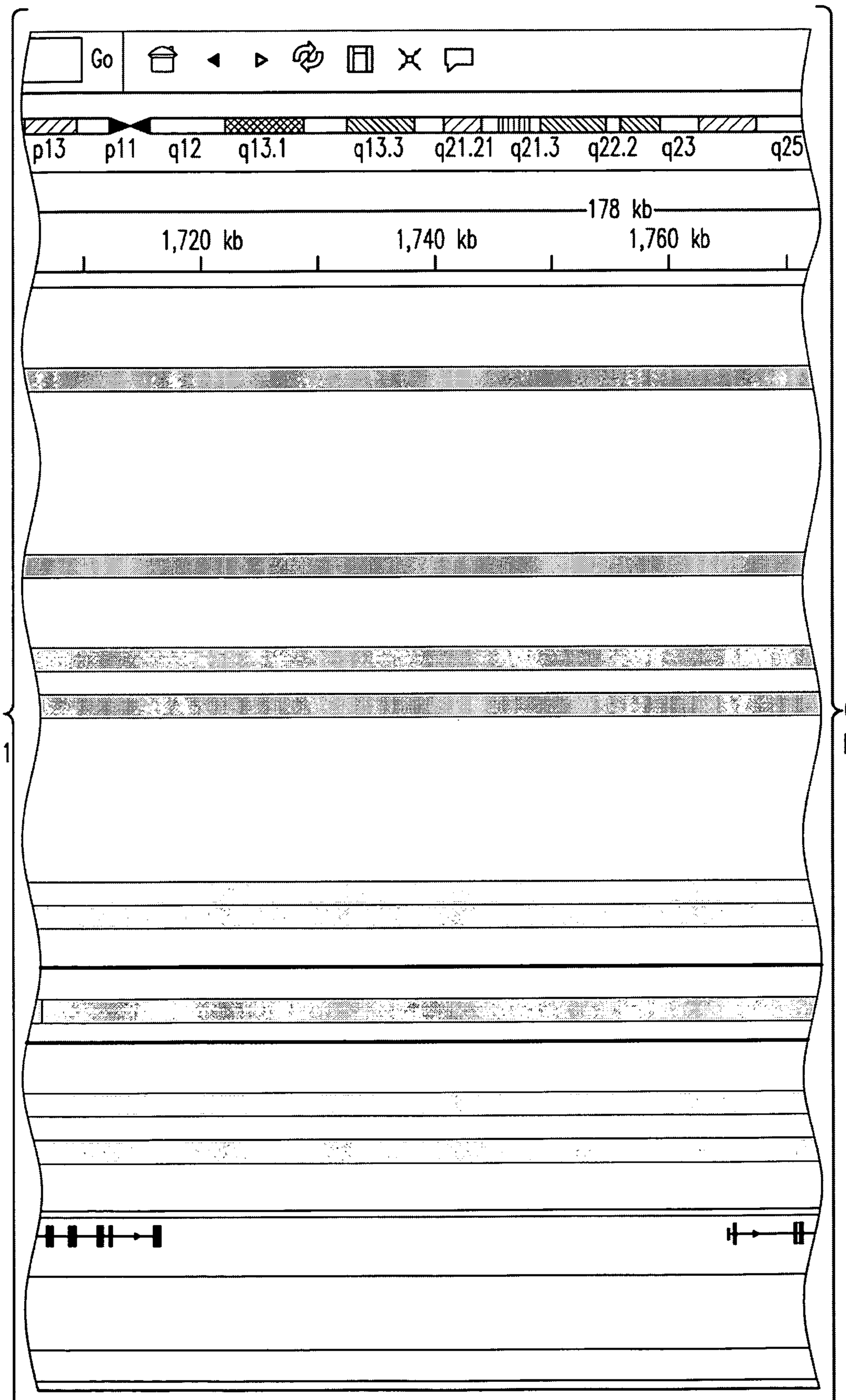

Go
p13
p11
q12
q13.1
q13.3
q21.21
q21.3
q22.2
q23
q25
178 kb
1,720 kb
1,740 kb
1,760 kb
Cont. From FIG.35-1
Cont. on FIG.35-3

FIG.35-2

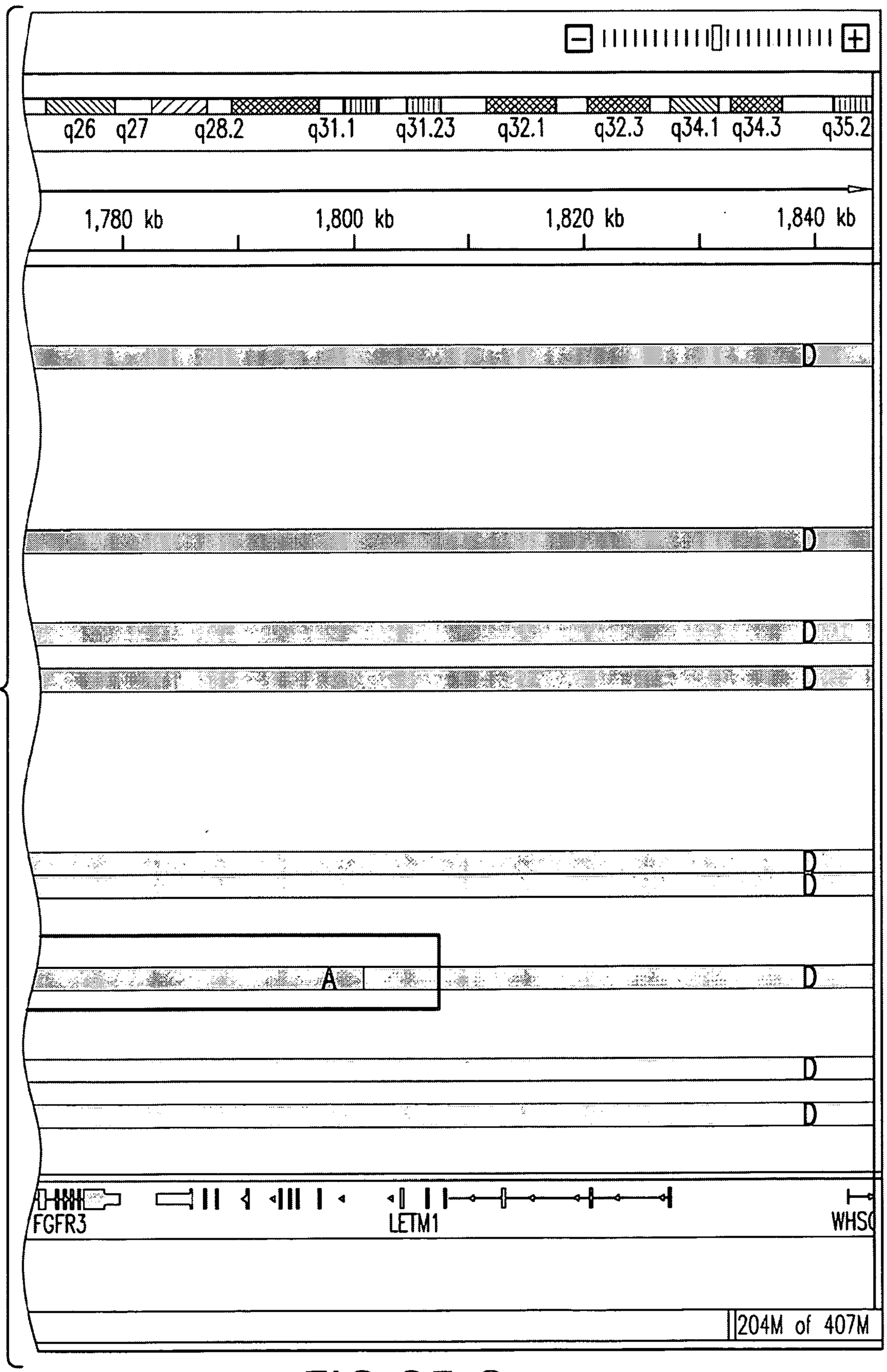

q26
q27
q28.2
q31.1
q31.23
q32.1
q32.3
q34.1
q34.3
q35.2
1,780 kb
1,800 kb
1,820 kb
1,840 kb
Cont.
From
FIG.35-2
A
FGFR3
LETM1
WHSC
204M of 407M

FIG.35-3

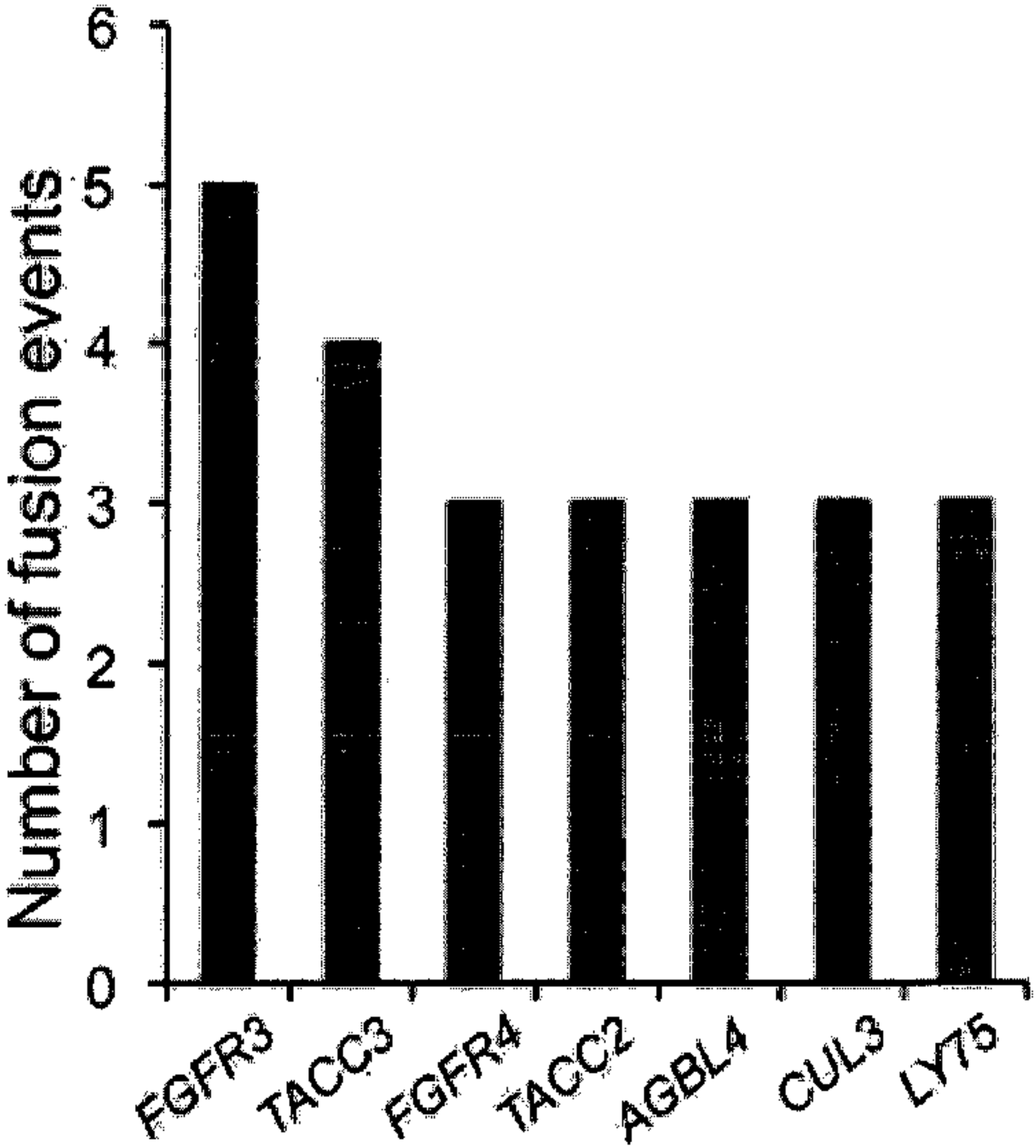

Number of fusion events
0
1
2
3
4
5
6
FGFR3
TACC3
FGFR4
TACC2
AGBL4
CUL3
LY75

FIG. 1A