(11) Application No. **AU 2005228446 B2**

(19) **AUSTRALIAN PATENT OFFICE**

(54) Title
Method to determine responsiveness of cancer to epidermal growth factor receptor targeting treatments

(51)^6 International Patent Classification(s)
C12Q 1/68 (2006.01) 20060101AFI2006080
C12Q 1/68 4BHAU
PCT/US2005/010645

(21) Application No: 2005228446 (22) Application Date: 2005 .03 .31

(87) WIPO No: WO05/094357

(30) Priority Data

(31) Number	(32) Date	(33) Country
60/565,753	2004 .04 .27	US
60/574,035	2004 .05 .25	US
60/558,218	2004 .03 .31	US
60/592,287	2004 .07 .29	US
60/577,916	2004 .06 .07	US
60/561,095	2004 .04 .09	US
60/565,985	2004 .04 .27	US

(43) Publication Date : 2005 .10 .13

(71) Applicant(s)
The General Hospital Corporation; Dana-Farber Cancer Institute, Inc.

(72) Inventor(s)
Johnson, Bruce E.; Settleman, Jeffrey E.; Paez, Juan Guillermo; Haber, Daniel A.; Janne, Pasi Antero; Bell, Daphne Winifred; Meyerson, Matthew; Sellers, William R.; Sordella, Raffaella; Lynch, Thomas J.

(74) Agent/Attorney
PIZZEYS, Level 14 ANZ Centre 324 Queen Street, Brisbane, QLD, 4001

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

(19) **World Intellectual Property Organization**
International Bureau

(43) **International Publication Date**
13 October 2005 (13.10.2005)

PCT

(10) **International Publication Number**
WO 2005/094357 A2

(51) **International Patent Classification:** **Not classified**

(21) **International Application Number:** PCT/US2005/010645

(22) **International Filing Date:** 31 March 2005 (31.03.2005)

(25) **Filing Language:** English

(26) **Publication Language:** English

(30) **Priority Data:**

60/558,218	31 March 2004 (31.03.2004)	US
60/561,095	9 April 2004 (09.04.2004)	US
60/565,753	27 April 2004 (27.04.2004)	US
60/565,985	27 April 2004 (27.04.2004)	US
60/574,035	25 May 2004 (25.05.2004)	US
60/577,916	7 June 2004 (07.06.2004)	US
60/592,287	29 July 2004 (29.07.2004)	US

(71) **Applicants** *(for all designated States except US)*: **THE GENERAL HOSPITAL CORPORATION** [US/US]; 55 Fruit Street, Boston, MA 02114 (US). **DANA-FARBER CANCER INSTITUTE, INC.** [US/US]; 44 Binney Street, Boston, MA 02115 (US).

(72) **Inventors; and**

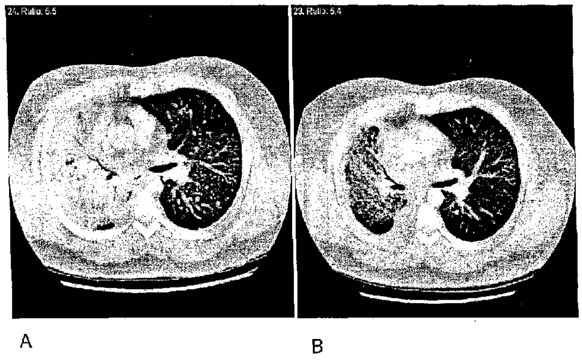
(75) **Inventors/Applicants** *(for US only)*: **BELL, Daphne, Winifred** [GB/US]; 7 Arizona Terrace, Apartment #3, Arlington, MA 02474 (US). **HABER, Daniel, A.** [US/US]; 34 Monadnock Rd., Chestnut Hill, MA 02467 (US). **JANNE, Pasi, Antero** [US/US]; 128 Roundwood Road, Newton, MA 02464 (US). **JOHNSON, Bruce, E.** [US/US]; 80 Seaver Street, Brookline, MA 02445 (US). **LYNCH, Thomas, J.** [US/US]; 131 Prince Street, Newton, MA 02465 (US). **MEYERSON, Matthew** [US/US]; 307 Independence Road, Concord, MA 01742 (US). **PAEZ, Juan, Guillermo** [AR/US]; 352 Riverway #22, Boston, MA 02115 (US). **SELLERS, William, R.** [US/US]; 80 Seaver Street, Brookline, MA 02445 (US). **SETTLEMAN, Jeffrey, E.** [US/US]; 140 Waverly Avenue, Newton, MA 02458 (US). **SORDELLA, Raffaella** [IT/US]; 38 Sunset Road, Bedford, MA 01730 (US).

(74) **Agent: RESNICK, David, S.**; Nixon Peabody LLP, 100 Summer Street, Boston, MA 02110-2131 (US).

(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI,

[Continued on next page]

(54) **Title:** METHOD TO DETERMINE RESPONSIVENESS OF CANCER TO EPIDERMAL GROWTH FACTOR RECEPTOR TARGETING TREATMENTS



(57) **Abstract:** The present invention is directed to a method for determining the responsiveness of cancer to an epidermal growth factor receptor (EGFR) treatment. In a preferred embodiment, the presence of at least one variance in the kinase domain of the erbB 1 gene confers sensitivity to the tyrosine kinase inhibitor gefitinib. Thus, a diagnostic assay for these mutations will allow for the administration of gefitinib, erlotinib and other tyrosine kinase inhibitors to those patients most likely to respond to the drug.

GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHOD TO DETERMINE RESPONSIVENESS OF CANCER TO EPIDERMAL GROWTH FACTOR RECEPTOR TARGETING TREATMENTS

CROSS REFERENCE TO RELATED APPLICATIONS

[001] The present application claims benefit under 35 U.S.C. § 119(e) of U.S. Provisional Application Serial No.: 60/558,218 filed March 31, 2004, U.S. Provisional Application Serial No.: 60/561,095 filed April 9, 2004, U.S. Provisional Application Serial No.: 60/565,753 filed April 27, 2004, U.S. Provisional Application Serial No.: 60/565,985 filed April 27, 2004, U.S. Provisional Application Serial No.: 60/574,035 filed May 25, 2004, U.S. Provisional Application Serial No.: 60/577,916 filed June 7, 2004 and U.S. Provisional Application Serial No.: 60/592,287 filed July 29, 2004, the contents of which are herein incorporated by reference in its entirety.

GOVERNMENT SUPPORT

[002] This invention was supported by the National Institutes for Health (NIH) Grant Nos. RO1 CA 092824, P50 CA 090578, PO1 95281, and 1K12CA87723-01, and the Government of the United States has certain rights thereto.

BACKGROUND

[003] Epithelial cell cancers, for example, prostate cancer, breast cancer, colon cancer, lung cancer, pancreatic cancer, ovarian cancer, cancer of the spleen, testicular cancer, cancer of the thymus, etc., are diseases characterized by abnormal, accelerated growth of epithelial cells. This accelerated growth initially causes a tumor to form. Eventually, metastasis to different organ sites can also occur. Although progress has been made in the diagnosis and treatment of various cancers, these diseases still result in significant mortality.

[004] Lung cancer remains the leading cause of cancer death in industrialized countries. Cancers that begin in the lungs are divided into two major types, non-small cell lung cancer and small cell lung cancer, depending on how the

cells appear under a microscope. Non-small cell lung cancer (squamous cell carcinoma, adenocarcinoma, and large cell carcinoma) generally spreads to other organs more slowly than does small cell lung cancer. About 75 percent of lung cancer cases are categorized as non-small cell lung cancer (e.g., adenocarcinomas), and the other 25 percent are small cell lung cancer. Non-small cell lung cancer (NSCLC) is the leading cause of cancer deaths in the United States, Japan and Western Europe. For patients with advanced disease, chemotherapy provides a modest benefit in survival, but at the cost of significant toxicity, underscoring the need for therapeutic agents that are specifically targeted to the critical genetic lesions that direct tumor growth (Schiller JH et al., N Engl J Med, 346: 92-98, 2002).

[005] Epidermal growth factor receptor (EGFR) is a 170 kilodalton (kDa) membrane-bound protein expressed on the surface of epithelial cells. EGFR is a member of the growth factor receptor family of protein tyrosine kinases, a class of cell cycle regulatory molecules. (W. J. Gullick et al., 1986, Cancer Res., 46:285-292). EGFR is activated when its ligand (either EGF or TGF-α) binds to the extracellular domain, resulting in autophosphorylation of the receptor's intracellular tyrosine kinase domain (S. Cohen et al., 1980, J. Biol. Chem., 255:4834-4842; A. B. Schreiber et al., 1983, J. Biol. Chem., 258:846-853).

[006] EGFR is the protein product of a growth promoting oncogene, erbB or ErbB1, that is but one member of a family, i.e., the ERBB family of protooncogenes, believed to play pivotal roles in the development and progression of many human cancers. In particular, increased expression of EGFR has been observed in breast, bladder, lung, head, neck and stomach cancer as well as glioblastomas. The ERBB family of oncogenes encodes four, structurally-related transmembrane receptors, namely, EGFR, HER-2/neu (erbB2), HER-3 (erbB3) and HER-4 (erbB4). Clinically, ERBB oncogene amplification and/or receptor overexpression in tumors have been reported to correlate with disease recurrence and poor patient prognosis, as well as with responsiveness in therapy. (L. Harris et al., 1999, Int. J. Biol. Markers, 14:8-15; and J. Mendelsohn and J. Baselga, 2000, Oncogene, 19:6550-6565).

[007] EGFR is composed of three principal domains, namely, the extracellular domain (ECD), which is glycosylated and contains the ligand-binding

pocket with two cysteine-rich regions; a short transmembrane domain, and an intracellular domain that has intrinsic tyrosine kinase activity. The transmembrane region joins the ligand-binding domain to the intracellular domain. Amino acid and DNA sequence analysis, as well as studies of nonglycosylated forms of EGFR, indicate that the protein backbone of EGFR has a mass of 132 kDa, with 1186 amino acid residues (A. L. Ullrich et al., 1984, Nature, 307:418-425; J. Downward et al., 1984, Nature, 307:521-527; C. R. Carlin et al., 1986, Mol. Cell. Biol., 6:257-264; and F. L. V. Mayes and M. D. Waterfield, 1984, The EMBO J., 3:531-537).

[008] The binding of EGF or TGF-α to EGFR activates a signal transduction pathway and results in cell proliferation. The dimerization, conformational changes and internalization of EGFR molecules function to transmit intracellular signals leading to cell growth regulation (G. Carpenter and S. Cohen, 1979, Ann. Rev. Biochem., 48:193-216). Genetic alterations that affect the regulation of growth factor receptor function, or lead to overexpression of receptor and/or ligand, result in cell proliferation. In addition, EGFR has been determined to play a role in cell differentiation, enhancement of cell motility, protein secretion, neovascularization, invasion, metastasis and resistance of cancer cells to chemotherapeutic agents and radiation. (M.-J. Oh et al., 2000, Clin. Cancer Res., 6:4760-4763).

[009] A variety of inhibitors of EGFR have been identified, including a number already undergoing clinical trials for treatment of various cancers. For a recent summary, see de Bono, J. S. and Rowinsky, E. K. (2002), "The ErbB Receptor Family: A Therapeutic Target For Cancer", *Trends in Molecular Medicine,* 8, S19-26.

[0010] A promising set of targets for therapeutic intervention in the treatment of cancer includes the members of the HER-kinase axis. They are frequently upregulated in solid epithelial tumors of, by way of example, the prostate, lung and breast, and are also upregulated in glioblastoma tumors. Epidermal growth factor receptor (EGFR) is a member of the HER-kinase axis, and has been the target of choice for the development of several different cancer therapies. EGFR tyrosine kinase inhibitors (EGFR-TKIs) are among these therapies, since the reversible phosphorylation of tyrosine residues is required for activation of the EGFR pathway.

In other words, EGFR-TKIs block a cell surface receptor responsible for triggering and/or maintaining the cell signaling pathway that induces tumor cell growth and division. Specifically, it is believed that these inhibitors interfere with the EGFR kinase domain, referred to as HER-1. Among the more promising EGFR-TKIs are three series of compounds: quinazolines, pyridopyrimidines and pyrrolopyrimidines.

[0011] Two of the more advanced compounds in clinical development include Gefitinib (compound ZD1839 developed by AstraZeneca UK Ltd.; available under the tradename IRESSA; hereinafter "IRESSA") and Erlotinib (compound OSI-774 developed by Genentech, Inc. and OSI Pharmaceuticals, Inc.; available under the tradename TARCEVA; hereinafter "TARCEVA"); both have generated encouraging clinical results. Conventional cancer treatment with both IRESSA and TARCEVA involves the daily, oral administration of no more than 500 mg of the respective compounds. In May, 2003, IRESSA became the first of these products to reach the United States market, when it was approved for the treatment of advanced non-small cell lung cancer patients.

[0012] IRESSA is an orally active quinazoline that functions by directly inhibiting tyrosine kinase phosphorylation on the EGFR molecule. It competes for the adenosine triphosphate (ATP) binding site, leading to suppression of the HER-kinase axis. The exact mechanism of the IRESSA response is not completely understood, however, studies suggest that the presence of EGFR is a necessary prerequisite for its action.

[0013] A significant limitation in using these compounds is that recipients thereof may develop a resistance to their therapeutic effects after they initially respond to therapy, or they may not respond to EGFR-TKIs to any measurable degree at all. In fact, only 10-15 percent of advanced non-small cell lung cancer patients respond to EGFR kinase inhibitors. Thus, a better understanding of the molecular mechanisms underlying sensitivity to IRESSA and TARCEVA would be extremely beneficial in targeting therapy to those individuals whom are most likely to benefit from such therapy.

[0014] There is a significant need in the art for a satisfactory treatment of cancer, and specifically epithelial cell cancers such as lung, ovarian, breast, brain, colon and prostate cancers, which incorporates the benefits of TKI therapy and overcoming the non-responsiveness exhibited by patients. Such a treatment could have a dramatic impact on the health of individuals, and especially older individuals, among whom cancer is especially common.

SUMMARY

[0015] Tyrosine kinase inhibitor (TKI) therapy such as gefitinib (IRESSA®) is not effective in the vast majority of individuals that are affected with the cancers noted above. The present inventors have surprisingly discovered that the presence of somatic mutations in the kinase domain of EGFR substantially increases sensitivity of the EGFR to TKI such as IRESSA, TARCEVA. For example less than 30% of patients having such cancer are susceptible to treatment by current TKIs, whereas greater than 50%, more preferably 60, 70, 80, 90 % of patients having a mutation in the EGFR kinase domain are susceptible. In addition, these mutations confer increased kinase activity of the EGFR. Thus, patients having these mutations will likely be responsive to current tyrosine kinase inhibitor (TKI) therapy, for example, gefitinib.

[0016] Accordingly, the present invention provides a novel method to determine the likelihood of effectiveness of an epidermal growth factor receptor (EGFR) targeting treatment in a human patient affected with cancer. The method comprises detecting the presence or absence of at least one nucleic acid variance in the kinase domain of the erbB1 gene of said patient relative to the wildtype erbB1 gene. The presence of at least one variance indicates that the EGFR targeting treatment is likely to be effective. Preferably, the nucleic acid variance increases the kinase activity of the EGFR. The patient can then be treated with an EGFR targeting treatment. In one embodiment of the present invention, the EGFR targeting treatment is a tyrosine kinase inhibitor. In a preferred embodiment, the tyrosine kinase inhibitor is an anilinoquinazoline. The anilinoquinazoline may be a synthetic anilinoquinazoline. Preferably, the synthetic anilinoquinazoline is either gefitinib or erlotinib. In another embodiment, the EGFR targeting treatment is an irreversible EGFR inhibitor, including 4-dimethylamino-but-2-enoic acid [4-(3-chloro-4-fluorophenylamino)-3-cyano-7-ethoxy-quinolin-6-yl]-amide ("EKB-569", sometimes also referred to as "EKI-569", see for example WO/2005/018677 and Torrance et al., Nature Medicine, vol. 6, No. 9, Sept. 2000, p. 1024) and/or HKI-272 or HKI-357 (Wyeth; see Greenberger et al., Proc. 11th NCI EORTC-AACR Symposium on New Drugs in Cancer Therapy, Clinical Cancer Res. Vol. 6 Supplement, Nov. 2000, ISSN 1078-0432; in Rabindran et al., Cancer Res. 64: 3958-3965 (2004); Holbro and Hynes, Ann. Rev. Pharm. Tox. 44:195-217 (2004); Tsou et al, j. Med. Chem. 2005, 48, 1107-1131; and Tejpar et al., J. Clin. Oncol. ASCO Annual Meeting Proc. Vol. 22, No. 14S: 3579 (2004)).

[0017] In one embodiment of the present invention, the EGFR is obtained from a biological sample from a patient with or at risk for developing cancer. The variance in the kinase domain of EGFR (or the erbB1 gene) effects the conformational structure of the ATP-binding pocket. Preferably, the variance in the kinase domain of EGFR is an in frame deletion or a substitution in exon 18, 19, 20 or 21.

[0018] In one embodiment, the in frame deletion is in exon 19 of EGFR (erbB1). The in frame deletion in exon 19 preferably comprises at deletion of at least amino acids leucine, arginine, glutamic acid and alanine, at codons 747, 748, 749, and 750. In one embodiment, the in-frame deletion comprises nucleotides 2235 to 2249 and deletes amino acids 746 to 750 (the sequence glutamic acid, leucine, arginine, glutamic acid, and alanine), see Table 2, Table S2, Figure 2B, Figure 4A, Figure 5, SEQ ID NO: 511, Figure 6C, and Figure 8C. In another embodiment, the in-frame deletion comprises nucleotides 2236 to 2250 and deletes amino acids 746 to 750, see Table S2, Figure 5, SEQ ID NO: 511, and Figure 6C. Alternatively, the in-frame deletion comprises nucleotides 2240 to 2251, see Table 2, Figure 2C, Figure 4A, Figure 5, SEQ ID NO: 511, or nucleotides 2240 to 2257, see Table 2, Table S3A, Figure 2A, Figure 4A, Figure 5, SEQ ID NO: 511, Figure 6C, and Figure 8E. Alternatively, the in-frame deletion comprises nucleotides 2239 to 2247 together with a substitution of cytosine for guanine at nucleotide 2248, see Table S3A and Figure 8D, or a deletion of nucleotides 2238 to 2255 together with a substitution of thymine for adenine at nucleotide 2237, see Table S3A and Figure 8F, or a deletion of nucleotides 2254 to 2277, see Table S2. Alternatively, the in-frame deletion comprises nucleotides 2239-2250delTTAAGAGAAGCA; 2251A>C, or 2240-2254delTAAGAGAAGCA, or 2257-2271delCCGAAAGCCAACAAG, as shown in Table S3B.

[0019] In another embodiment, the substitution is in exon 21 of EGFR. The substitution in exon 21 comprises at least one amino acid. In one embodiment, the substitution in exon 21 comprises a substitution of a guanine for a thymine at nucleotide 2573, see Figure 4A and Figure 5, SEQ ID NO: 511. This substitution results in an amino acid substitution, where the wildtype Leucine is replaced with an Arginine at amino acid 858, see Figure 5, Table 2, Table S2, Table S3A, Figure 2D, Figure 6A, figure 8B, and SEQ ID NO: 512. Alternatively, the substitution in exon 21 comprises a substitution of an adenine for a thymine at nucleotide 2582, see Figure 4A and Figure 5, SEQ ID NO: 511. This substitution results in an amino acid substitution, where the wildtype Leucine is replaced with a Glutamine at amino acid 861, see Figure 5, Table 2, Figure 2E, Table S3B, and SEQ ID NO: 512.

[0020] The substitution may also be in exon 18 of EGFR. In one embodiment, the substitution is in exon 18 is a thymine for a guanine at nucleotide 2155, see Figure 4A and Figure 5, SEQ ID NO: 511. This substitution results in an amino acid substitution, where the wildtype Glycine is substituted with a Cysteine at codon 719, see Figure 5, SEQ ID NO: 512. In another embodiment, the substitution in exon 18 is an adenine for a guanine at nucleotide 2155 resulting in an amino acid substitution, where the wildtype Glycine is substituted for a Serine at codon 719, see Table S2, Figure 6B, Figure 8A, Figure 5, SEQ ID NO: 511 and 512.

[0021] In another embodiment, the substitution is an insertion of guanine, guanine and thymine (GGT) after nucleotide 2316 and before nucleotide 2317 of SEQ ID NO: 511 (2316_2317 ins GGT). This can also be described as an insertion of valine (V) at amino acid 772 (P772_H733 insV). Other mutations are shown in Table S3B and include, for example, and insertion of CAACCCGG after nucleotide 2309 and before nucleotide 2310 of SEQ ID NO 511 and an insertion of GCGTGGACA after nucleotide 2311 and before nucleotide 2312 of SEQ ID NO 511. The substitution may also be in exon 20 and in one embodiment is a substitution of AA for GG at nucleotides 2334 and 2335, see Table S3B.

[0022] In summary, in preferred embodiments, the nucleic acid variance of the erbB1 gene is a substitution of a thymine for a guanine or an adenine for a guanine at nucleotide 2155 of SEQ ID NO 511, a deletion of nucleotides 2235 to 2249, 2240 to 2251, 2240 to 2257, 2236 to 2250, 2254 to 2277, or 2236 to 2244 of SEQ ID NO 511, an insertion of nucleotides guanine, guanine, and thymine (GGT) after nucleotide 2316 and before nucleotide 2317 of SEQ ID NO 511, and a substitution of a guanine for a thymine at nucleotide 2573 or an adenine for a thymine at nucleotide 2582 of SEQ ID NO 511.

[0023] The detection of the presence or absence of at least one nucleic acid variance can be determined by amplifying a segment of nucleic acid encoding the receptor. The segment to be amplified is 1000 nucleotides in length, preferably, 500 nucleotides in length, and most preferably 100 nucleotides in length or less. The segment to be amplified can include a plurality of variances.

[0024] In another embodiment, the detection of the presence or absence of at least one variance provides for contacting EGFR nucleic acid containing a variance site with at least one nucleic acid probe. The probe preferentially hybridizes with a nucleic acid sequence including a variance site and containing complementary nucleotide bases at the variance site under selective hybridization conditions. Hybridization can be detected with a detectable label.

[0025] In yet another embodiment, the detection of the presence or absence of at least one variance comprises sequencing at least one nucleic acid sequence and comparing the obtained sequence with the known erbB1 nucleic acid sequence. Alternatively, the presence or absence of at least one variance comprises mass spectrometric determination of at least one nucleic acid sequence.

[0026] In a preferred embodiment, the detection of the presence or absence of at least one nucleic acid variance comprises performing a polymerase chain reaction (PCR). The erbB1 nucleic acid sequence containing the hypothetical variance is amplified and the nucleotide sequence of the amplified nucleic acid is determined. Determining the nucleotide sequence of the amplified nucleic acid comprises sequencing at least one nucleic acid segment. Alternatively, amplification products can analyzed by using any method capable of separating the amplification products according to their size, including automated and manual gel electrophoresis and the like.

[0027] Alternatively, the detection of the presence or absence of at least one variance comprises determining the haplotype of a plurality of variances in a gene.

[0028] In another embodiment, the presence or absence of an EGFR variance can be detected by analyzing the erbB1 gene product (protein). In this embodiment, a probe that specifically binds to a variant EGFR is utilized. In a preferred embodiment, the probe is an antibody that preferentially binds to a variant EGFR. The presence of a variant EGFR predicts the likelihood of effectiveness of an EGFR targeting treatment. Alternatively, the probe may be an antibody fragment, chimeric antibody, humanized antibody or an aptamer.

[0029] The present invention further provides a probe which specifically binds under selective binding conditions to a nucleic acid sequence comprising at least one nucleic acid variance in the EGFR gene (erbB1). In one embodiment, the variance is a mutation in the kinase domain of erbB1 that confers a structural change in the ATP-binding pocket.

[0030] The probe of the present invention may comprise a nucleic acid sequence of about 500 nucleotide bases, preferably about 100 nucleotides bases, and most preferably about 50 or about 25 nucleotide bases or fewer in length. The probe may be composed of DNA, RNA, or peptide nucleic acid (PNA). Furthermore, the probe may contain a detectable label, such as, for example, a fluorescent or enzymatic label.

[0031] The present invention additionally provides a novel method to determine the likelihood of effectiveness of an epidermal growth factor receptor (EGFR) targeting treatment in a patient affected with cancer. The method comprises determining the kinase activity of the EGFR in a biological sample from a patient. An increase in kinase activity following stimulation with an EGFR ligand, compared to a normal control, indicates that the EGFR targeting treatment is likely to be effective.

[0032] The present invention further provides a novel method for treating a patient affected with or at risk for developing cancer. The method involves determining whether the kinase domain of the EGFR of a patient contains at least one nucleic acid variance. Preferably, the EGFR is located at the site of the tumor or cancer and the nucleic acid variance is somatic. The presence of such a variance indicates that an EGFR targeted treatment will be effective. If the variance is present, the tyrosine kinase inhibitor is administered to the patient.

[0033] As above, the tyrosine kinase inhibitor administered to an identified patient may be an anilinoquinazoline or an irreversible tyrosine kinase inhibitor, such as for example, EKB-569, HKI-272 and/or HKI-357 (Wyeth). Preferably, the anilinoquinazoline is a synthetic anilinoquinazoline and most preferably the synthetic anilinoquinazoline is gefitinib and erlotinib.

[0034] The cancer to be treated by the methods of the present invention include, for example, but are not limited to, gastrointestinal cancer, prostate cancer, ovarian cancer, breast cancer, head and neck cancer, lung cancer, non-small cell lung cancer, cancer of the nervous system, kidney cancer, retina cancer, skin cancer, liver cancer, pancreatic cancer, genital-urinary cancer and bladder cancer. In a preferred embodiment, the cancer is non-small cell lung cancer.

[0035] A kit for implementing the PCR methods of the present invention is also encompassed. The kit includes at least one degenerate primer pair designed to anneal to nucleic acid regions bordering the genes that encode for the ATP-binding pocket of the EGFR kinase domain. Additionally, the kit contains the products and reagents required to carry out PCR amplification, and instructions.

[0036] In a preferred embodiment, the primer pairs contained within the kit are selected from the group consisting of SEQ ID NO: 505, SEQ ID NO: 506, SEQ ID NO: 507, and SEQ ID NO: 508. Also preferred are the primers listed in Table 6 and 7 in the examples.

[0037] In yet another embodiment, the present invention discloses a method for selecting a compound that inhibits the catalytic kinase activity of a variant epidermal growth factor receptor (EGFR). As a first step, a variant EGFR is contacted with a potential compound. The resultant kinase activity of the variant

EGFR is then detected and a compound is selected that inhibits the kinase activity of the variant EGFR. In one embodiment, the variant EGFR is contained within a cell. The method can also be used to select a compound that inhibits the kinase activity of a variant EGFR having a secondary mutation in the kinase domain that confers resistance to a TKI, e.g., gefitinib or erlotinib.

[0038] In one embodiment, the variant EGFR is labeled. In another embodiment, the EGFR is bound to a solid support. In a preferred embodiment, the solid support is a protein chip.

[0039] In yet another embodiment of the present invention, a pharmaceutical composition that inhibits the catalytic kinase activity of a variant epidermal growth factor receptor (EGFR) is disclosed. The compound that inhibits the catalytic kinase activity of a variant EGFR is selected from the group consisting of an antibody, antibody fragment, small molecule, peptide, protein, antisense nucleic acid, ribozyme, PNA, siRNA, oligonucleotide aptamer, and peptide aptamer.

[0040] A method for treating a patient having an EGFR mediated disease is also disclosed. In accordance with the method, the patient is administered the pharmaceutical composition that inhibits the catalytic kinase activity of a variant epidermal growth factor receptor (EGFR).

[0041] In one embodiment, the EGFR mediated disease is cancer. In a preferred embodiment, the cancer is of epithelial origin. For example, the cancer is gastrointestinal cancer, prostate cancer, ovarian cancer, breast cancer, head and neck cancer, lung cancer, non-small cell lung cancer, cancer of the nervous system, kidney cancer, retina cancer, skin cancer, liver cancer, pancreatic cancer, genital-urinary cancer and bladder cancer. In a preferred embodiment, the cancer is non-small cell lung cancer.

[0042] In another embodiment, a method for predicting the acquisition of secondary mutations (or selecting for mutations) in the kinase domain of the erbB1 gene is disclosed. A cell expressing a variant form of the erbB1 gene is contacted with an effective, yet sub-lethal dose of a tyrosine kinase inhibitor. Cells that are resistant to a growth arrest effect of the tyrosine kinase inhibitor are selected and the erbB1 nucleic acid is analyzed for the presence of additional mutations in the erbB1 kinase domain. In one embodiment, the cell is *in vitro*. In another embodiment, the cell is obtained from a transgenic animal. In one embodiment, the transgenic animal is a mouse. In this mouse model, cells to be studied are obtained from a tumor biopsy. Cells containing a secondary mutation in the erbB1 kinase domain selected by the present invention can be used in the above methods to select a compound that inhibits the kinase activity of the variant EGFR having a secondary mutation in the kinase domain.

[0043] In an alternative embodiment for predicting the acquisition of secondary mutations in the kinase domain of the erbB1 gene, cells expressing a variant form of the erbB1 gene are first contacted with an effective amount of a mutagenizing agent. The mutagenizing is, for example, ethyl methanesulfonate (EMS), N-ethyl-N-nitrosourea (ENU), N-methyl-N-nitrosourea (MNU), phocarbaxine hydrochloride (Prc), methyl methanesulfonate (MeMS), chlorambucil (Chl), melphalan, porcarbazine hydrochloride, cyclophosphamide (Cp), diethyl sulfate (Et_2SO_4), acrylamide monomer (AA), triethylene melamin (TEM), nitrogen mustard, vincristine, dimethylnitrosamine, N-methyl-N'-nitro-Nitrosoguanidine (MNNG), 7,12 dimethylbenz(a)anthracene (DMBA), ethylene oxide, hexamethylphosphoramide, bisulfan, or ethyl methanesulforate (EtMs). The cell is then contacted with an effective, yet sub-lethal dose of a tyrosine kinase inhibitor. Cells that are resistant to a growth arrest effect of the tyrosine kinase inhibitor are selected and the erbB1 nucleic acid is analyzed for the presence of additional mutations in the erbB1 kinase domain.

BRIEF DESCRIPTION OF THE FIGURES

[0044] Figures 1A-1B show a representative illustration of Gefitinib response in refractory non-small cell lung cancer (NSCLC). Chest CT scan of case 6 (Table 1) , demonstrating (Figure 1A) a large mass in the right lung before treatment with gefitinib, and (Figure 1B) marked improvement six weeks after Gefitinib was initiated.

[0045] Figure 2 shows EGFR mutations in Gefitinib-responsive tumors.

[0046] Figures 2A -2C show nucleotide sequence of the *EGFR* gene in tumor specimens with heterozygous in-frame deletions within the kinase domain (double peaks) (SEQ ID NOS 643-654, respectively, in order of appearance). Tracings in both sense and antisense directions are shown to demonstrate the two breakpoints of the deletion; wild-type nucleotide sequence is shown in capital letters, and the mutant sequence is in lowercase letters. The 5' breakpoint of the delL747-T751insS mutation is preceded by a T to C substitution that does not alter the encoded amino acid.

[0047] Figure 2D and Figure 2E show heterozygous missense mutations (arrows) resulting in amino acid substitutions within the tyrosine kinase domain (SEQ ID NOS 656 & 658). The double peaks represent two nucleotides at the site of heterozygous mutations. For comparison, the corresponding wild-type sequence is also shown (SEQ ID NOS 655 & 657).

[0048] Figure 2F is a schematic representation of dimerized EGFR molecules bound by the EGF ligand. The extracellular domain (containing two receptor ligand [L]-domains and a furin-like domain), transmembrane region, and the cytoplasmic domain (containing the catalytic kinase domain) are highlighted. The position of tyrosine^{1068} (Y-1068), a site of autophosphorylation used as a marker of receptor activation, is indicated, along with downstream effectors activated by EGFR autophosphorylation (STAT3, MAP Kinase (MAPK), and AKT). The location of tumor-associated mutations, all within the tyrosine kinase domain, is shown.

[0049] Figure 3 demonstrates enhanced EGF-dependent activation of mutant EGFR and increased sensitivity of mutant EGFR to Gefitinib.

[0050] Figure 3A shows a time course of ligand-induced activation of the delL747-P753insS and L858R mutants, compared with wild type EGFR, following addition of EGF to serum starved cells. EGFR autophosphorylation is used as a marker of receptor activation, using Western blotting with an antibody that specifically recognizes the phosphorylated tyrosine ^{1068} residue of EGFR (left panel), compared with the total levels of EGFR expressed in Cos-7 cells (control; right panel). Autophosphorylation of EGFR is measured at intervals following addition of EGF (10 ng/ml).

[0051] Figure 3B is a graphical representation of EGF-induction of wild-type and mutant receptor phosphorylation (see panel A). Autoradiographs from three independent experiments were quantified using the NIH image software; intensity of EGFR phosphorylation is normalized to total protein expression, and shown as percent activation of the receptor, with standard deviation.

[0052] Figure 3C shows a dose-dependent inhibition of EGFR activation by Gefitinib. Autophosphorylation of EGFR tyrosine ^{1068} is demonstrated by Western blotting analysis of Cos-7 cells expressing wild-type or mutant receptors, and stimulated with 100 ng/ml of EGF for 30 min. Cells were untreated (U) or pretreated for 3 hrs with increasing concentrations of Gefitinib as shown (left panel). Total amounts of EGFR protein expressed are shown as control (right panel).

[0053] Figure 3D shows the quantification of results from two experiments described for panel 3C (NIH image software). Concentrations of phosphorylated EGFR were normalized to protein expression levels and expressed as percent activation of the receptor.

[0054] Figure 4 demonstrates clustering of mutations at critical sites within the ATP- binding pocket of EGFR.

[0055] Figure 4A shows the position of overlapping in-frame deletions in exon 19 and missense mutations in exon 21 of the *EGFR* gene, in multiple cases of NSCLC (SEQ ID NOS 495-504 (DNA)). Partial nucleotide sequence is shown for each exon, with deletions marked by dashed lines and missense mutations highlighted and underlined; the wild-type *EGFR* nucleotide and amino acid sequences are shown (SEQ ID NOS 493 & 494 (DNA) & 509-510 (amino acid)).

[0056] Figure 4B shows the tridimensional structure of the EGFR ATP cleft flanked by the amino (N) and carboxy (C) lobes of the kinase domain (coordinates derived from PDB 1M14, and displayed using Cn3D software). The inhibitor, representing Gefitinib, is pictured occupying the ATP cleft. The locations of the two missense mutations are shown, within the activating loop of the kinase; the three in-frame deletions are all present within another loop, which flanks the ATP cleft.

[0057] Figure 4C is a close-up of the EGFR kinase domain, showing the critical amino acid residues implicated in binding to either ATP or to the inhibitor. Specifically, 4-anilinoquinazoline compounds such as gefitinib inhibit catalysis by occupying the ATP-binding site, where they form hydrogen bonds with methionine^{793} (M793) and cysteine^{775} (C775) residues, whereas their anilino ring is close to methionine^{766} (M766), lysine^{745} (K745),and leucine^{788} (L788) residues . In-frame deletions within the loop that is targeted by mutations are predicted to alter the position of these amino acids relative to the inhibitor. Mutated residues are shown within the activation loop of the tyrosine kinase.

[0058] Figure 5 shows the nucleotide and amino acid sequence of the erbB1 gene. The amino acids are depicted as single letters, known to those of skill in the art. Nucleotide variances in the kinase domain are highlighted by patient number, see Table 2. SEQ ID NO: 511 includes nucleotides 1 through 3633. SEQ ID NO: 512 includes amino acids 1 through 1210.

[0059] Figures 6A-6C: Sequence alignment of selected regions within the EGFR and B-Raf kinase domains. Depiction of EGFR mutations in human NSCLC. EGFR (gb:X00588;) mutations in NSCLC tumors are highlighted in gray. B-Raf (gb:M95712) mutations in multiple tumor types (5) are highlighted in black. Asterisks denote residues conserved between EGFR and B-Raf. FIG. 6A depicts L858R mutations in the activation loop (SEQ ID NOS 477-479). FIG. 6B depicts the G719S mutant in the P-loop (SEQ ID NOS 480-482). FIG. 6C depicts deletion mutants in EGFR exon 19 (SEQ ID NOS 483-489).

[0060] Figure 7: Positions of missense mutations G719S and L858R and the Del-1 deletion in the three-dimensional structure of the EGFR kinase domain. The activation loop is shown in yellow, the P-loop is in blue and the C-lobe and N-lobe are as indicated. The residues targeted by mutation or deletion are highlighted in red. The Del-1 mutation targets the residues ELREA in codons 746 to 750. The mutations are located in highly conserved regions within kinases and are found in the p-loop and activation loop, which surround the region where ATP and also gefitinib and erlotinib are predicted to bind.

[0061] Figures 8A - 8F. Representative chromatograms of EGFR DNA from normal tissue and from tumor tissues. The locations of the identified mutations are as follows. Fig. 8A depicts the Exon 18 Kinase domain P loop (SEQ ID NOS 659-660). Fig. 8B depicts the Exon 21 Kinase domain A-loop (SEQ ID NOS 661-

662). Fig. 8C depicts the Exon 19 Kinase domain Del-1 (SEQ ID NOS 663-665). Fig. 8D depicts the Exon 19 Kinase domain Del-3 (SEQ ID NOS 666-668). Fig. 8E depicts the Exon 19 Kinase domain Del-4 (SEQ ID NOS 669-671). Fig. 8F depicts the Exon 19 Kinase domain Del-5 (SEQ ID NOS 672-674).

[0062] FIG. 9: Sequence alignment of the EGFR and BCR-ABL polypeptides and the location of residues conferring a drug resistant phenotype. The EGFR polypeptide (SEQ ID NO:492) encoded by the nucleotide sequence disclosed in GenBank accno. NM_005228 and the BCR-ABL polypeptide (SEQ ID NO:491) encoded by the nucleotide sequence disclosed in GenBank accno. M14752 are aligned and conserved residues are shaded. BCR-ABL mutations conferring resistance to the tyrosine kinase inhibitor imatinib (STI571, Glivec/Gleevec) are denoted by asterisks.

[0063] Figure 10 shows the decision making process for patient with metastatic NSCLC undergoing EGFR testing.

[0064] Figure 11 shows a diagram of EGFR exons 18-24 (not to scale). Arrows deptict the location of identified mutations. Astericks denote the number of patients with mutations at each location. The blow-up diagram depicts the overlap of the exon 19 deletions, and the number of patients (n) with each deletion. Note that these are the results are not meant to be inclusive of all the EGFR mutations to date.

DETAILED DESCRIPTION

[0065] The present invention provides a novel method to determine the likelihood of effectiveness of an epidermal growth factor receptor (EGFR) targeting treatment in a patient affected with cancer. The method comprises detecting the presence or absence of at least one nucleic acid variance in the kinase domain of the erbB1 gene of said patient. The presence of at least one variance indicates that the EGFR targeting treatment is likely to be effective. Preferably, the nucleic acid variance increases the kinase activity of the EGFR. The patient can then be treated with an EGFR targeting treatment. In one embodiment of the present invention, the EGFR targeting treatment is a tyrosine kinase inhibitor. In a preferred embodiment, the tyrosine kinase inhibitor is an anilinoquinazoline. The anilinoquinazoline may be a synthetic anilinoquinazoline. Preferably, the synthetic anilinoquinazoline is either gefitinib or erlotinib.

Definitions:

[0066] The terms "ErbB1", "epidermal growth factor receptor" and "EGFR" are used interchangeably herein and refer to native sequence EGFR as disclosed, for example, in Carpenter et al. Ann. Rev. Biochem. 56:881-914 (1987), including variants thereof (e.g. a deletion mutant EGFR as in Humphrey et al. PNAS (USA) 87:4207-4211 (1990)). erbB1 refers to the gene encoding the EGFR protein product.

[0067] The term "kinase activity increasing nucleic acid variance" as used herein refers to a variance (i.e. mutation) in the nucleotide sequence of a gene that results in an increased kinase activity. The increased kinase activity is a direct result of the variance in the nucleic acid and is associated with the protein for which the gene encodes.

[0068] The term "drug" or "compound" as used herein refers to a chemical entity or biological product, or combination of chemical entities or biological products, administered to a person to treat or prevent or control a disease or condition. The chemical entity or biological product is preferably, but not necessarily a low molecular weight compound, but may also be a larger compound, for example, an oligomer of nucleic acids, amino acids, or carbohydrates including without limitation proteins, oligonucleotides, ribozymes, DNAzymes, glycoproteins, siRNAs, lipoproteins, aptamers, and modifications and combinations thereof.

[0069] The term "genotype" in the context of this invention refers to the particular allelic form of a gene, which can be defined by the particular nucleotide(s) present in a nucleic acid sequence at a particular site(s).

[0070] The terms "variant form of a gene", "form of a gene", or "allele" refer to one specific form of a gene in a population, the specific form differing from other forms of the same gene in the sequence of at least one, and frequently more than one, variant sites within the sequence of the gene. The sequences at these variant sites that differ between different alleles of the gene are termed "gene sequence variances" or "variances" or "variants". Other terms known in the art to be equivalent include mutation and polymorphism, although mutation is often used to refer to an allele associated with a deleterious phenotype. In preferred aspects of this invention, the variances are selected from the group consisting of the variances listed in the variance tables herein.

[0071] In the context of this invention, the term "probe" refers to a molecule which can detectably distinguish between target molecules differing in structure. Detection can be accomplished in a variety of different ways depending on the type of probe used and the type of target molecule. Thus, for example, detection may be based on discrimination of activity levels of the target molecule, but preferably is based on detection of specific binding. Examples of such specific binding include antibody binding and nucleic acid probe hybridization. Thus, for example, probes can include enzyme substrates, antibodies and antibody fragments, and preferably nucleic acid hybridization probes.

[0072] As used herein, the terms "effective" and "effectiveness" includes both pharmacological effectiveness and physiological safety. Pharmacological effectiveness refers to the ability of the treatment to result in a desired biological effect in the patient. Physiological safety refers to the level of toxicity, or other adverse physiological effects at the cellular, organ and/or organism level (often referred to as side-effects) resulting from administration of the treatment. "Less effective" means that the treatment results in a therapeutically significant lower level of pharmacological effectiveness and/or a therapeutically greater level of adverse physiological effects.

[0073] The term "primer", as used herein, refers to an oligonucleotide which is capable of acting as a point of initiation of polynucleotide synthesis along a complementary strand when placed under conditions in which synthesis of a primer extension product which is complementary to a polynucleotide is catalyzed. Such conditions include the presence of four different nucleotide triphosphates or nucleoside analogs and one or more agents for polymerization such as DNA polymerase and/or reverse transcriptase, in an appropriate buffer ("buffer" includes substituents which are cofactors, or which affect pH, ionic strength, etc.), and at a suitable temperature. A primer must be sufficiently long to prime the synthesis of extension products in the presence of an agent for polymerase. A typical primer contains at least about 5 nucleotides in length of a sequence substantially complementary to the target sequence, but somewhat longer primers are preferred. Usually primers contain about 15-26 nucleotides, but longer primers may also be employed.

[0074] A primer will always contain a sequence substantially complementary to the target sequence, that is the specific sequence to be amplified, to which it can anneal. A primer may, optionally, also comprise a promoter sequence. The term "promoter sequence" defines a single strand of a nucleic acid sequence that is specifically recognized by an RNA polymerase that binds to a recognized sequence and initiates the process of transcription by which an RNA transcript is produced. In principle, any promoter sequence may be employed for which there is a known and available polymerase that is capable of recognizing the initiation sequence. Known and useful promoters are those that are recognized by certain bacteriophage polymerases, such as bacteriophage T3, T7 or SP6.

[0075] A "microarray" is a linear or two-dimensional array of preferably discrete regions, each having a defined area, formed on the surface of a solid support. The density of the discrete regions on a microarray is determined by the total numbers of target polynucleotides to be detected on the surface of a single solid phase support, preferably at least about 50/cm^2, more preferably at least about 100/cm^2, even more preferably at least about 500/cm^2, and still more preferably at least about 1,000/cm^2. As used herein, a DNA microarray is an array of oligonucleotide primers placed on a chip or other surfaces used to amplify or clone target polynucleotides. Since the position of each particular group of primers in the array is known, the identities of the target polynucleotides can be determined based on their binding to a particular position in the microarray.

[0076] The term "label" refers to a composition capable of producing a detectable signal indicative of the presence of the target polynucleotide in an assay sample. Suitable labels include radioisotopes, nucleotide chromophores, enzymes, substrates, fluorescent molecules, chemiluminescent moieties, magnetic particles, bioluminescent moieties, and the like. As such, a label is any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means.

[0077] The term "support" refers to conventional supports such as beads, particles, dipsticks, fibers, filters, membranes and silane or silicate supports such as glass slides.

[0078] The term "amplify" is used in the broad sense to mean creating an amplification product which may include, for example, additional target molecules, or target-like molecules or molecules complementary to the target molecule, which molecules are created by virtue of the presence of the target molecule in the sample. In the situation where the target is a nucleic acid, an amplification product can be made enzymatically with DNA or RNA polymerases or reverse transcriptases.

[0079] As used herein, a "biological sample" refers to a sample of tissue or fluid isolated from an individual, including but not limited to, for example, blood, plasma, serum, tumor biopsy, urine, stool, sputum, spinal fluid, pleural fluid, nipple aspirates, lymph fluid, the external sections of the skin, respiratory, intestinal, and genitourinary tracts, tears, saliva, milk, cells (including but not limited to blood cells), tumors, organs, and also samples of *in vitro* cell culture constituent. In a preferred embodiment, the sample is from a resection, bronchoscopic biopsy, or core needle biopsy of a primary or metastatic tumor, or a cellblock from pleural fluid. In addition, fine needle aspirate samples are used. Samples may be either paraffin-embedded or frozen tissue.

[0080] The term "antibody" is meant to be an immunoglobulin protein that is capable of binding an antigen. Antibody as used herein is meant to include antibody fragments, e.g. F(ab')2, Fab', Fab, capable of binding the antigen or antigenic fragment of interest. Preferably, the binding of the antibody to the antigen inhibits the activity of a variant form of EGFR.

[0081] The term "humanized antibody" is used herein to describe complete antibody molecules, i.e. composed of two complete light chains and two complete heavy chains, as well as antibodies consisting only of antibody fragments, e.g. Fab, Fab', F (ab') 2, and Fv, wherein the CDRs are derived from a non-human source and the remaining portion of the Ig molecule or fragment thereof is derived from a human antibody, preferably produced from a nucleic acid sequence encoding a human antibody.

[0082] The terms "human antibody" and "humanized antibody" are used herein to describe an antibody of which all portions of the antibody molecule are derived from a nucleic acid sequence encoding a human antibody. Such human antibodies are most desirable for use in antibody therapies, as such antibodies would elicit little or no immune response in the human patient.

[0083] The term "chimeric antibody" is used herein to describe an antibody molecule as well as antibody fragments, as described above in the definition of the term "humanized antibody." The term "chimeric antibody" encompasses humanized antibodies. Chimeric antibodies have at least one portion of a heavy or light chain amino acid sequence derived from a first mammalian species and another portion of the heavy or light chain amino acid sequence derived from a second, different mammalian species.

[0084] Preferably, the variable region is derived from a non-human mammalian species and the constant region is derived from a human species. Specifically, the chimeric antibody is preferably produced from a 9 nucleotide sequence from a non-human mammal encoding a variable region and a nucleotide sequence from a human encoding a constant region of an antibody.

[0085] Table 2 is a partial list of DNA sequence variances in the kinase domain of erbB1 relevant to the methods described in the present invention. These variances were identified by the inventors in studies of biological samples from patients with NSCLC who responded to gefitinib and patients with no exposure to gefitinb.

[0086] Nucleic acid molecules can be isolated from a particular biological sample using any of a number of procedures, which are well-known in the art, the particular isolation procedure chosen being appropriate for the particular biological sample. For example, freeze-thaw and alkaline lysis procedures can be useful for obtaining nucleic acid molecules from solid materials; heat and alkaline lysis procedures can be useful for obtaining nucleic acid molecules from urine; and proteinase K extraction can be used to obtain nucleic acid from blood (Rolff, A et al. PCR: Clinical Diagnostics and Research, Springer (1994).

Detection Methods

[0087] Determining the presence or absence of a particular variance or plurality of variances in the kinase domain of the erbB1 gene in a patient with or at risk for developing cancer can be performed in a variety of ways. Such tests are commonly performed using DNA or RNA collected from biological samples, e.g., tissue biopsies, urine, stool, sputum, blood, cells, tissue scrapings, breast aspirates or other cellular materials, and can be performed by a variety of methods including, but not limited to, PCR, hybridization with allele-specific probes, enzymatic mutation detection, chemical cleavage of mismatches, mass spectrometry or DNA sequencing, including minisequencing. In particular embodiments, hybridization with allele specific probes can be conducted in two formats: (1) allele specific oligonucleotides bound to a solid phase (glass, silicon, nylon membranes) and the labeled sample in solution, as in many DNA chip applications, or (2) bound sample (often cloned DNA or PCR amplified DNA) and labeled oligonucleotides in solution (either allele specific or short so as to allow sequencing by hybridization). Diagnostic tests may involve a panel of variances, often on a solid support, which enables the simultaneous determination of more than one variance.

[0088] In another aspect, determining the presence of at least one kinase activity increasing nucleic acid variance in the erbB1 gene may entail a haplotyping test. Methods of determining haplotypes are known to those of skill in the art, as for example, in WO 00/04194.

[0089] Preferably, the determination of the presence or absence of a kinase activity increasing nucleic acid variance involves determining the sequence of the variance site or sites by methods such as polymerase chain reaction (PCR). Alternatively, the determination of the presence or absence of a kinase activity increasing nucleic acid variance may encompass chain terminating DNA sequencing or minisequencing, oligonucleotide hybridization or mass spectrometry.

[0090] The methods of the present invention may be used to predict the likelihood of effectiveness (or lack of effectiveness) of an EGFR targeting treatment in a patient affected with or at risk for developing cancer. Preferably, cancers include cancer of epithelial origin, including, but are not limited to, gastrointestinal cancer, prostate cancer, ovarian cancer, breast cancer, head and neck cancer, lung cancer, non-small cell lung cancer, cancer of the nervous system, kidney cancer, retina cancer, skin cancer, liver cancer, pancreatic cancer, genital-urinary cancer and bladder cancer. In a preferred embodiment, the cancer is non-small cell lung cancer.

[0091] The present invention generally concerns the identification of variances in the kinase domain of the erbB1 gene which are indicative of the effectiveness of an EGFR targeting treatment in a patient with or at risk for developing cancer. Additionally, the identification of specific variances in the kinase domain of EGFR, in effect, can be used as a diagnostic or prognostic test. For example, the presence of at least one variance in the kinase domain of erbB1 indicates that a patient will likely benefit from treatment with an EGFR targeting compound, such as, for example, a tyrosine kinase inhibitor.

[0092] Methods for diagnostic tests are well known in the art and disclosed in patent application WO 00/04194, incorporated herein by reference. In an exemplary method, the diagnostic test comprises amplifying a segment of DNA or RNA (generally after converting the RNA to cDNA) spanning one or more known variances in the kinase domain of the erbB1 gene sequence. This amplified segment is then sequenced and/or subjected to polyacrylamide gel electrophoresis in order to identify nucleic acid variances in the amplified segment.

PCR

[0093] In one embodiment, the invention provides a method of screening for variants in the kinase domain of the erbB1 gene in a test biological sample by PCR or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran, et al., 1988. Science 241: 1077-1080; and Nakazawa, et al., 1994. Proc. Natl. Acad. Sci. USA 91: 360-364), the latter of which can be particularly useful for detecting point mutations in the EGFR-gene (see, Abravaya, et al., 1995. Nucl. Acids Res. 23: 675-682). The method comprises the steps of designing degenerate primers for amplifying the target sequence, the primers corresponding to one or more conserved regions of the gene, amplifying reaction with the primers using, as a template, a DNA or cDNA obtained from a test biological sample and analyzing the PCR products. Comparison of the PCR products of the test biological sample to a control sample indicates variances in the test biological sample. The change can be either and absence or presence of a nucleic acid variance in the test biological sample.

[0094] Alternative amplification methods include: self sustained sequence replication (see, Guatelli, et al., 1990. Proc. Natl. Acad. Sci. USA 87: 1874-1 878), transcriptional amplification system (see, Kwoh, et al., 1989. Proc. Natl. Acad. Sci. USA 86: 1173-1177); Qb Replicase (see, Lizardi, et al, 1988. BioTechnology 6: 1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

[0095] Primers useful according to the present invention are designed using amino acid sequences of the protein or nucleic acid sequences of the kinase domain of the erbB1 gene as a guide, e.g. SEQ ID NO: 493, SEQ ID NO: 494, SEQ ID NO: 509, and SEQ ID NO: 510. The primers are designed in the homologous regions of the gene wherein at least two regions of homology are separated by a divergent region of variable sequence, the sequence being variable either in length or nucleic acid sequence.

[0096] For example, the identical or highly, homologous, preferably at least 80%-85% more preferably at least 90-99% homologous amino acid sequence of at least about 6, preferably at least 8-10 consecutive amino acids. Most preferably, the amino acid sequence is 100% identical. Forward and reverse primers are designed based upon the maintenance of codon degeneracy and the representation of the various amino acids at a given position among the known gene family members. Degree of homology as referred to herein is based upon analysis of an amino acid sequence using a standard sequence comparison software, such as protein-BLAST using the default settings (http://www.ncbi.nlm.nih.gov/BLAST/).

[0097] Table 3 below represents the usage of degenerate codes and their standard symbols:

	T	C	A	G
T	TTT Phe (F)	TCT Ser (S)	TAT Tyr (Y)	TGT Cys (C)
	TTC "	TCC "	TAC	TGC
	TTA Leu (L)	TCA "	TAA Ter	TGA Ter
	TTG "	TCG "	TAG Ter	TGG Trp (W)
C	CTT Leu (L)	CCT Pro (P)	CAT His (H)	CGT Arg (R)
	CTC "	CCC "	CAC "	CGC "
	CTA "	CCA "	CAA Gln (Q)	CGA "
	CTG "	CCG "	CAG "	CGG "
A	ATT Ile (I)	ACT Thr (T)	AAT Asn (N)	AGT Ser (S)
	ATC "	ACC "	AAC "	AGC "
	ATA "	ACA "	AAA Lys (K)	AGA Arg (R)
	ATG Met (M)	ACG "	AAG "	AGG "
G	GTT Val (V)	GCT Ala (A)	GAT Asp (D)	GGT Gly (G)
	GTC "	GCC "	GAC "	GGC "
	GTA "	GCA "	GAA Glu (E)	GGA "
	GTG "	GCG "	GAG "	GGG "

[0098] Preferably any 6-fold degenerate codons such as L, R and S are avoided since in practice they will introduce higher than 6-fold degeneracy. In the case of L, TTR and CTN are compromised YTN (8-fold degeneracy), in the case of R, CGN and AGR compromises at MGN (8-fold degeneracy), and finally S, TCN and AGY which can be compromised to WSN (16-fold degeneracy). In all three cases on 6 of these will match the target sequence. To avoid this loss of specificity, it is preferable to avoid these regions, or to make two populations, each with the alternative degenerate codon, e.g. for S include TCN in one pool, and AGY in the other.

[0099] Primers may be designed using a number of available computer programs, including, but not limited to Oligo Analyzer3.0; Oligo Calculator; NetPrimer; Methprimer; Primer3; WebPrimer; PrimerFinder; Primer9;

Oligo2002; Pride or GenomePride; Oligos; and Codehop. Detailed information about these programs can be obtained, for example, from www.molbiol.net.

[00100] Primers may be labeled using labels known to one skilled in the art. Such labels include, but are not limited to radioactive, fluorescent, dye, and enzymatic labels.

[00101] Analysis of amplification products can be performed using any method capable of separating the amplification products according to their size, including automated and manual gel electrophoresis, mass spectrometry, and the like.

[00102] Alternatively, the amplification products can be separated using sequence differences, using SSCP, DGGE, TGGE, chemical cleavage or restriction fragment polymorphisms as well as hybridization to, for example, a nucleic acid arrays.

[00103] The methods of nucleic acid isolation, amplification and analysis are routine for one skilled in the art and examples of protocols can be found, for example, in the Molecular Cloning: A Laboratory Manual (3-Volume Set) Ed. Joseph Sambrook, David W. Russel, and Joe Sambrook, Cold Spring Harbor Laboratory; 3rd edition (January 15, 2001), ISBN: 0879695773. Particularly useful protocol source for methods used in PCR amplification is PCR (Basics: From Background to Bench) by M. J. McPherson, S. G. Møller, R. Beynon, C. Howe, Springer Verlag; 1st edition (October 15, 2000), ISBN: 0387916008.

[00104] Preferably, exons 19 and 21 of human EGFR are amplified by the polymerase chain reaction (PCR) using the following primers: Exon19 sense primer, 5'- GCAATATCAGCCTTAGGTGCGGCTC-3' (SEQ ID NO: 505); Exon 19 antisense primer, 5'-CATAGAA AGTGAACATTTAGGATGTG-3' (SEQ ID NO: 506); Exon 21 sense primer, 5'-CTAACGTTCG CCAGCCATAAGTCC-3' (SEQ ID NO: 507); and Exon21 antisense primer, 5'- GCTGCGAGCTCACCCAG AATGTCTGG-3' (SEQ ID NO: 508).

[00105] In an alternative embodiment, mutations in a EGFR gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined

by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, e.g., U.S. Patent No. 5,493,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

[00106] Other methods for detecting mutations in the EGFR gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes. See, e.g., Myers, et al., 1985. Science 230: 1242. In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type EGFR sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, e.g., Cotton, et al., 1988. Proc. Natl. Acad. Sci. USA 85: 4397; Saleeba, et al., 1992. Methods Enzymol. 2 17: 286-295. In an embodiment, the control DNA or RNA can be labeled for detection.

[00107] In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in EGFR cDNAs obtained from samples of cells. For example, the mutY enzyme of E. coli cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches. See, e.g., Hsu, et al., 1994. Carcinogenesis 15: 1657-1662. According to an exemplary embodiment, a probe based on a mutant EGFR sequence, e.g., a DEL-1 through DEL-5, G719S, G857V, L883S or L858R EGFR sequence, is hybridized to a cDNA or

other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, e.g., U.S. Patent No. 5,459,039.

[00108] In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in EGFR genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids. See, e.g., Orita, et al., 1989. Proc. Natl. Acad. Sci. USA: 86: 2766; Cotton, 1993. Mutat. Res. 285: 125-144; Hayashi, 1992. Genet. Anal. Tech. Appl. 9: 73-79. Single-stranded DNA fragments of sample and control EGFR nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. See, e.g., Keen, et al., 1991. Trends Genet. 7: 5.

[00109] In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE). See, e.g., Myers, et al., 1985. Nature 313: 495. When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA. See, e.g., Rosenbaum and Reissner, 1987. Biophys. Chem. 265: 12753.

[00110] Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA

under conditions that permit hybridization only if a perfect match is found. See, e.g., Saiki, et al., 1986. Nature 324: 163; Saiki, et al., 1989. Proc. Natl. Acad. Sci. USA 86: 6230. Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

[00111] Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization; see, e.g., Gibbs, et al., 1989. Nucl. Acids Res. 17: 2437-2448) or at the extreme 3'-terminus of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (see, e.g., Prossner, 1993. Tibtech. 11: 238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection. See, e.g., Gasparini, et al., 1992. Mol. Cell Probes 6: 1. It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification. See, e.g., Barany, 1991. Proc. Natl. Acad. Sci. USA 88: 189. In such cases, ligation will occur only if there is a perfect match at the 3'-terminus of the 5' sequence, making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

Solid Support and Probe

[00112] In an alternative embodiment, the detection of the presence or absence of the at least one nucleic acid variance involves contacting a nucleic acid sequence corresponding to the desired region of the erbB1 gene, identified above, with a probe. The probe is able to distinguish a particular form of the gene or the presence or a particular variance or variances, e.g., by differential binding or hybridization. Thus, exemplary probes include nucleic acid hybridization probes, peptide nucleic acid probes, nucleotide-containing probes which also contain at least one nucleotide analog, and antibodies, e.g., monoclonal antibodies, and other probes as discussed herein. Those skilled in the art are familiar with the preparation of probes

with particular specificities. Those skilled in the art will recognize that a variety of variables can be adjusted to optimize the discrimination between two variant forms of a gene, including changes in salt concentration, temperature, pH and addition of various compounds that affect the differential affinity of GC vs. AT base pairs, such as tetramethyl ammonium chloride. (See Current Protocols in Molecular Biology by F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, K. Struhl and V. B. Chanda (Editors), John Wiley & Sons.)

[00113] Thus, in preferred embodiments, the detection of the presence or absence of the at least one variance involves contacting a nucleic acid sequence which includes at least one variance site with a probe, preferably a nucleic acid probe, where the probe preferentially hybridizes with a form of the nucleic acid sequence containing a complementary base at the variance site as compared to hybridization to a form of the nucleic acid sequence having a non-complementary base at the variance site, where the hybridization is carried out under selective hybridization conditions. Such a nucleic acid hybridization probe may span two or more variance sites. Unless otherwise specified, a nucleic acid probe can include one or more nucleic acid analogs, labels or other substituents or moieties so long as the base-pairing function is retained.

[00114] The probe may be designed to bind to, for example, at least three continuous nucleotides on both sides of the deleted region of SEQ ID NO: 495, SEQ ID NO: 497, or SEQ ID NO: 499. Such probes, when hybridized under the appropriate conditions, will bind to the variant form of EGFR, but will not bind to the wildtype EGFR.

[00115] Such hybridization probes are well known in the art (see, e.g., Sambrook et al., Eds., (most recent edition), Molecular Cloning: A Laboratory Manual, (third edition, 2001), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.). Stringent hybridization conditions will typically include salt concentrations of less than about 1M, more usually less than about 500 mM and preferably less than about 200 mM. Hybridization temperatures can be as low as 5°C, but are typically greater than 22°C, more typically greater than about 30°C, and preferably in excess of about 37°C. Longer fragments may require higher

hybridization temperatures for specific hybridization. Other factors may affect the stringency of hybridization, including base composition and length of the complementary strands, presence of organic solvents and extent of base mismatching; the combination of parameters used is more important than the absolute measure of any one alone. Other hybridization conditions which may be controlled include buffer type and concentration, solution pH, presence and concentration of blocking reagents (e.g., repeat sequences, Cot1 DNA, blocking protein solutions) to decrease background binding, detergent type(s) and concentrations, molecules such as polymers which increase the relative concentration of the polynucleotides, metal ion(s) and their concentration(s), chelator(s) and their concentrations, and other conditions known or discoverable in the art. Formulas may be used to predict the optimal melting temperature for a perfectly complementary sequence for a given probe, but true melting temperatures for a probe under a set of hybridization conditions must be determined empirically. Also, a probe may be tested against its exact complement to determine a precise melting temperature under a given set of condition as described in Sambrook et al, "Molecular Cloning," 3rd edition, Cold Spring Harbor Laboratory Press, 2001. Hybridization temperatures can be systematically altered for a given hybridization solution using a support associated with target polynucleotides until a temperature range is identified which permits detection of binding of a detectable probe at the level of stringency desired, either at high stringency where only target polynucleotides with a high degree of complementarity hybridize, or at lower stringency where additional target polynucleotides having regions of complementarity with the probe detectably hybridize above the background level provided from nonspecific binding to noncomplementary target polynucleotides and to the support. When hybridization is performed with potential target polynucleotides on a support under a given set of conditions, the support is then washed under increasing conditions of stringency (typically lowered salt concentration and/or increased temperature, but other conditions may be altered) until background binding is lowered to the point where distinct positive signals may be seen. This can be monitored in progress using a Geiger counter where the probe is radiolabeled, radiographically, using a fluorescent imager, or by other means of detecting probe binding. The support is not allowed to dry during such procedures, or the probe may become irreversibly bound even to background locations. Where a probe produces undesirable background or false positives, blocking reagents are employed, or different regions of the probe or different probes are used until positive signals can be distinguished from background. Once conditions are found that provide satisfactory signal above background, the target polynucleotides providing a positive signal are isolated and further characterized. The isolated polynucleotides can be sequenced; the sequence can be compared to databank entries or known sequences; where necessary, full-length clones can be obtained by techniques known in the art; and the polynucleotides can be expressed using suitable vectors and hosts to determine if the polynucleotide identified encodes a protein having similar activity to that from which the probe polynucleotide was derived. The probes can be from 10-50 nucleotides. However, musch oarger probes can also be employed, e.g., 50-500 nucleotides or larger.

Solid Phase Support

[00116] The solid phase support of the present invention can be of any solid materials and structures suitable for supporting nucleotide hybridization and synthesis. Preferably, the solid phase support comprises at least one substantially rigid surface on which oligonucleotides or oligonucleotide primers can be immobilized. The solid phase support can be made of, for example, glass, synthetic polymer, plastic, hard non-mesh nylon or ceramic. Other suitable solid support materials are known and readily available to those of skill in the art. The size of the solid support can be any of the standard microarray sizes, useful for DNA microarray technology, and the size may be tailored to fit the particular machine being used to conduct a reaction of the invention. Methods and materials for derivatization of solid phase supports for the purpose of immobilizing oligonucleotides are known to those skill in the art and described in, for example, U.S. Pat. No. 5,919,523, the disclosure of which is incorporated herein by reference.

[00117] The solid support can be provided in or be part of a fluid containing vessel. For example, the solid support can be placed in a chamber with sides that create a seal along the edge of the solid support so as to contain the polymerase chain reaction (PCR) on the support. In a specific example the chamber can have walls on each side of a rectangular support to ensure that the PCR mixture remains on the support and also to make the entire surface useful for providing the primers.

[00118] The oligonucleotide or oligonucleotide primers of the invention are affixed, immobilized, provided, and/or applied to the surface of the solid support using any available means to fix, immobilize, provide and/or apply the oligonucleotides at a particular location on the solid support. For example, photolithography (Affymetrix, Santa Clara, Calif.) can be used to apply the oligonucleotide primers at particular position on a chip or solid support, as described in the U.S. Pat. Nos. 5,919,523, 5,837,832, 5,831,070, and 5,770,722, which are incorporated herein by reference. The oligonucleotide primers may also be applied to a solid support as described in Brown and Shalon, U.S. Pat. No. 5,807,522 (1998). Additionally, the primers may be applied to a solid support using a robotic system, such as one manufactured by Genetic MicroSystems (Woburn, Mass.), GeneMachines (San Carlos, Calif.) or Cartesian Technologies (Irvine, Calif.).

[00119] In one aspect of the invention, solid phase amplification of target polynucleotides from a biological sample is performed, wherein multiple groups of oligonucleotide primers are immobilized on a solid phase support. In a preferred embodiment, the primers within a group comprises at least a first set of primers that are identical in sequence and are complementary to a defined sequence of the target polynucleotide, capable of hybridizing to the target polynucleotide under appropriate conditions, and suitable as initial primers for nucleic acid synthesis (i.e., chain elongation or extension). Selected primers covering a particular region of the reference sequence are immobilized, as a group, onto a solid support at a discrete location. Preferably, the distance between groups is greater than the resolution of detection means to be used for detecting the amplified products. In a preferred embodiment, the primers are immobilized to form a microarray or chip that can be processed and analyzed via automated, processing. The immobilized primers are used

for solid phase amplification of target polynucleotides under conditions suitable for a nucleic acid amplification means. In this manner, the presence or absence of a variety of potential variances in the kinase domain of the erbB1 gene can be determined in one assay.

[00120] A population of target polynucleotides isolated from a healthy individual can used as a control in determining whether a biological source has at least one kinase activity increasing variance in the kinase domain of the erb1 gene. Alternatively, target polynucleotides isolated from healthy tissue of the same individual may be used as a control as above.

[00121] An in situ-type PCR reactions on the microarrays can be conducted essentially as described in e.g. Embretson et al, Nature 362:359-362 (1993); Gosden et al, BioTechniques 15(1):78-80 (1993); Heniford et al Nuc. Acid Res. 21(14):3159-3166 (1993); Long et al, Histochemistry 99:151-162 (1993); Nuovo et al, PCR Methods and Applications 2(4):305-312 (1993); Patterson et al Science 260:976-979 (1993).

[00122] Alternatively, variances in the kinase domain of erbB1 can be determined by solid phase techniques without performing PCR on the support. A plurality of oligonucleotide probes, each containing a distinct variance in the kinase domain of erbB1, in duplicate, triplicate or quadruplicate, may be bound to the solid phase support. The presence or absence of variances in the test biological sample may be detected by selective hybridization techniques, known to those of skill in the art and described above.

Mass Spectrometry

[00123] In another embodiment, the presence or absence of kinase activity increasing nucleic acid variances in the kinase domain of the erbB1 gene are determined using mass spectrometry. To obtain an appropriate quantity of nucleic acid molecules on which to perform mass spectrometry, amplification may be necessary. Examples of appropriate amplification procedures for use in the invention include: cloning (Sambrook et al., Molecular Cloning: A Laboratory Manual, 3rd Edition, Cold Spring Harbor Laboratory Press, 2001), polymerase chain reaction

(PCR) (C. R. Newton and A. Graham, PCR, BIOS Publishers, 1994), ligase chain reaction (LCR) (Wiedmann, M., et al., (1994) PCR Methods Appl. Vol. 3, Pp. 57-64; F. Barnay Proc. Natl. Acad. Sci USA 88, 189-93 (1991), strand displacement amplification (SDA) (G. Terrance Walker et al., Nucleic Acids Res. 22, 2670-77 (1994)) and variations such as RT-PCR (Higuchi, et al., Bio/Technology 11:1026-1030 (1993)), allele-specific amplification (ASA) and transcription based processes.

[00124] To facilitate mass spectrometric analysis, a nucleic acid molecule containing a nucleic acid sequence to be detected can be immobilized to a solid support. Examples of appropriate solid supports include beads (e.g. silica gel, controlled pore glass, magnetic, Sephadex/Sepharose, cellulose), flat surfaces or chips (e.g. glass fiber filters, glass surfaces, metal surface (steel, gold, silver, aluminum, copper and silicon), capillaries, plastic (e.g. polyethylene, polypropylene, polyamide, polyvinylidenedifluoride membranes or microtiter plates)); or pins or combs made from similar materials comprising beads or flat surfaces or beads placed into pits in flat surfaces such as wafers (e.g. silicon wafers).

[00125] Immobilization can be accomplished, for example, based on hybridization between a capture nucleic acid sequence, which has already been immobilized to the support and a complementary nucleic acid sequence, which is also contained within the nucleic acid molecule containing the nucleic acid sequence to be detected. So that hybridization between the complementary nucleic acid molecules is not hindered by the support, the capture nucleic acid can include a spacer region of at least about five nucleotides in length between the solid support and the capture nucleic acid sequence. The duplex formed will be cleaved under the influence of the laser pulse and desorption can be initiated. The solid support-bound base sequence can be presented through natural oligoribo- or oligodeoxyribonucleotide as well as analogs (e.g. thio-modified phosphodiester or phosphotriester backbone) or employing oligonucleotide mimetics such as PNA analogs (see e.g. Nielsen et al., Science, 254, 1497 (1991)) which render the base sequence less susceptible to enzymatic degradation and hence increases overall stability of the solid support-bound capture base sequence.

[00126] Prior to mass spectrometric analysis, it may be useful to "condition" nucleic acid molecules, for example to decrease the laser energy required for volatilization and/or to minimize fragmentation. Conditioning is preferably performed while a target detection site is immobilized. An example of conditioning is modification of the phosphodiester backbone of the nucleic acid molecule (e.g. cation exchange), which can be useful for eliminating peak broadening due to a heterogeneity in the cations bound per nucleotide unit. Contacting a nucleic acid molecule with an alkylating agent such as alkyliodide, iodoacetamide, β-iodoethanol, 2,3-epoxy-1-propanol, the monothio phosphodiester bonds of a nucleic acid molecule can be transformed into a phosphotriester bond. Likewise, phosphodiester bonds may be transformed to uncharged derivatives employing trialkylsilyl chlorides. Further conditioning involves incorporating nucleotides which reduce sensitivity for depurination (fragmentation during MS) such as N7- or N9-deazapurine nucleotides, or RNA building blocks or using oligonucleotide triesters or incorporating phosphorothioate functions which are alkylated or employing oligonucleotide mimetics such as PNA.

[00127] For certain applications, it may be useful to simultaneously detect more than one (mutated) loci on a particular captured nucleic acid fragment (on one spot of an array) or it may be useful to perform parallel processing by using oligonucleotide or oligonucleotide mimetic arrays on various solid supports. "Multiplexing" can be achieved by several different methodologies. For example, several mutations can be simultaneously detected on one target sequence by employing corresponding detector (probe) molecules (e.g. oligonucleotides or oligonucleotide mimetics). However, the molecular weight differences between the detector oligonucleotides D1, D2 and D3 must be large enough so that simultaneous detection (multiplexing) is possible. This can be achieved either by the sequence itself (composition or length) or by the introduction of mass-modifying functionalities M1-M3 into the detector oligonucleotide.

[00128] Preferred mass spectrometer formats for use in the invention are matrix assisted laser desorption ionization (MALDI), electrospray (ES), ion cyclotron resonance (ICR) and Fourier Transform. Methods of performing mass spectrometry

are known to those of skill in the art and are further described in Methods of Enzymology, Vol. 193:"Mass Spectrometry" (J. A. McCloskey, editor), 1990, Academic Press, New York.

Sequencing

[00129] In other preferred embodiments, determining the presence or absence of the at least one kinase activity increasing nucleic acid variance involves sequencing at least one nucleic acid sequence. The sequencing involves the sequencing of a portion or portions of the kinase domain of erbB1 which includes at least one variance site, and may include a plurality of such sites. Preferably, the portion is 500 nucleotides or less in length, more preferably 100 nucleotides or less, and most preferably 45 nucleotides or less in length. Such sequencing can be carried out by various methods recognized by those skilled in the art, including use of dideoxy termination methods (e.g., using dye-labeled dideoxy nucleotides), minisequencing, and the use of mass spectrometric methods.

Immunodetection

[00130] In one embodiment, determining the presence or absence of the at least one kinase activity increasing nucleic acid variance involves determining the activation state of downstream targets of EGFR.

[00131] The inventors of the present application have compared the phosphorylation status of the major downstream targets of EGFR. For example, the EGF-induced activation of Erk1 and Erk2, via Ras, of Akt via PLCγ/PI3K, and of STAT3 and STAT5 via JAK2, has been examined. Erk1 and Erk2, via Ras, Akt via PLCγ/PI3K, and STAT3 and STAT5 via JAK2 are essential downstream pathways mediating oncogenic effects of EGFR (R. N. Jorissen et al., *Exp. Cell Res.* **284**, 31 (2003)).

[00132] The inventors of the present application have shown that EGF-induced Erk activation is indistinguishable among cells expressing wild-type EGFR or either of the two activating EGFR mutants.

[00133] In contrast, phosphorylation of both Akt and STAT5 was substantially elevated in cells expressing either of the mutant EGFRs. Increased phosphorylation of STAT3 was similarly observed in cells expressing mutant EGFRs. Thus, the selective EGF-induced autophosphorylation of C-terminal tyrosine residues within EGFR mutants is well correlated with the selective activation of downstream signaling pathways.

[00134] In one embodiment of the present application, the presence of EGFR mutations can be determined using immunological techniques well known in the art, e.g., antibody techniques such as immunohistochemistry, immunocytochemistry, FACS scanning, immunoblotting, radioimmunoassays, western blotting, immunoprecipitation, enzyme-linked immunosorbant assays (ELISA), and derivative techniques that make use of antibodies directed against activated downstream targets of EGFR. Examples of such targets include, for example, phosphorylated STAT3, phosphorylated STAT5, and phosphorylated Akt. Using phospho-specific antibodies, the activation status of STAT3, STAT5, and Akt can be determined. Activation of STAT3, STAT5, and Akt are useful as a diagnostic indicator of activating EGFR mutations.

[00135] In one embodiment of the present invention, the presence of activated (phosphorylated) STAT5, STAT3, or Akt indicates that an EGFR targeting treatment is likely to be effective.

[00136] The invention provides a method of screening for variants in the kinase domain of the erbB1 gene in a test biological sample by immunohistochemical or immunocytochemical methods.

[00137] Immunohistochemistry ("IHC") and immunocytochemistry ("ICC") techniques, for example, may be used. IHC is the application of immunochemistry to tissue sections, whereas ICC is the application of immunochemistry to cells or tissue imprints after they have undergone specific cytological preparations such as, for example, liquid-based preparations. Immunochemistry is a family of techniques based on the use of a specific antibody, wherein antibodies are used to specifically target molecules inside or on the surface of cells. The antibody typically contains a marker that will undergo a biochemical reaction, and thereby experience a change color, upon

encountering the targeted molecules. In some instances, signal amplification may be integrated into the particular protocol, wherein a secondary antibody, that includes the marker stain, follows the application of a primary specific antibody.

[00138] Immunoshistochemical assays are known to those of skill in the art (e.g., see Jalkanen, et al., J. Cell. Biol. 101:976-985 (1985); Jalkanen, et al., J. Cell. Biol. 105:3087-3096 (1987).

[00139] Antibodies, polyclonal or monoclonal, can be purchased from a variety of commercial suppliers, or may be manufactured using well-known methods, e. g., as described in Harlow et al., Antibodies: A Laboratory Manual, 2nd Ed; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1988). In general, examples of antibodies useful in the present invention include anti-phospho-STAT3, anti-phospho-STAT5, and anti-phospho-Akt antibodies. Such antibodies can be purchased, for example, from Upstate Biotechnology (Lake Placid, NY), New England Biolabs (Beverly, MA), NeoMarkers (Fremont, CA)

[00140] Typically, for immunohistochemistry, tissue sections are obtained from a patient and fixed by a suitable fixing agent such as alcohol, acetone, and paraformaldehyde, to which is reacted an antibody. Conventional methods for immunohistochemistry are described in Harlow and Lane (eds) (1988) In "Antibodies A Laboratory Manual", Cold Spring Harbor Press, Cold Spring Harbor, New York; Ausbel et al (eds) (1987), in Current Protocols In Molecular Biology, John Wiley and Sons (New York, NY). Biological samples appropriate for such detection assays include, but are not limited to, cells, tissue biopsy, whole blood, plasma, serum, sputum, cerebrospinal fluid, breast aspirates, pleural fluid, urine and the like.

[00141] For direct labeling techniques, a labeled antibody is utilized. For indirect labeling techniques, the sample is further reacted with a labeled substance.

[00142] Alternatively, immunocytochemistry may be utilized. In general, cells are obtained from a patient and fixed by a suitable fixing agent such as alcohol, acetone, and paraformaldehyde, to which is reacted an antibody. Methods of immunocytological staining of human samples is known to those of skill in the art and described, for example, in Brauer et al., 2001 (FASEB J, 15, 2689- 2701), Smith-Swintosky et al., 1997.

[00143] Immunological methods of the present invention are advantageous because they require only small quantities of biological material. Such methods may be done at the cellular level and thereby necessitate a minimum of one cell. Preferably, several cells are obtained from a patient affected with or at risk for developing cancer and assayed according to the methods of the present invention.

Other Diagnostic Methods

[00144] An agent for detecting mutant EGFR protein is an antibody capable of binding to mutant EGFR protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, F_{ab} or $F_{(ab)2}$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect mutant EGFR mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of mutant EGFR mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of mutant EGFR protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of mutant EGFR genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of mutant EGFR protein include introducing into a subject a labeled anti- mutant EGFR protein antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

[00145] In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject.

[00146] In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting mutant EGFR protein, mRNA, or genomic DNA, such that the presence of mutant EGFR protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of mutant EGFR protein, mRNA or genomic DNA in the control sample with the presence of mutant EGFR protein, mRNA or genomic DNA in the test sample.

[00147] In a different embodiment, the diagnostic assay is for mutant EGFR activity. In a specific embodiment, the mutant EGFR activity is a tyrosine kinase activity. One such diagnostic assay is for detecting EGFR-mediated phosphorylation of at least one EGFR substrate. Levels of EGFR activity can be assayed for, *e.g.*, various mutant EGFR polypeptides, various tissues containing mutant EGFR, biopsies from cancer tissues suspected of having at least one mutant EGFR, and the like. Comparisons of the levels of EGFR activity in these various cells, tissues, or extracts of the same, can optionally be made. In one embodiment, high levels of EGFR activity in cancerous tissue is diagnostic for cancers that may be susceptible to treatments with one or more tyrosine kinase inhibitor. In related embodiments, EGFR activity levels can be determined between treated and untreated biopsy samples, cell lines, transgenic animals, or extracts from any of these, to determine the effect of a given treatment on mutant EGFR activity as compared to an untreated control.

Method of Treating a Patient

[00148] In one embodiment, the invention provides a method for selecting a treatment for a patient affected by or at risk for developing cancer by determining the presence or absence of at least one kinase activity increasing nucleic acid variance in the kinase domain of the erbB1 gene. In another embodiment, the variance is a

plurality of variances, whereby a plurality may include variances from one, two, three or more gene loci.

[00149] In certain embodiments, the presence of the at least one variance is indicative that the treatment will be effective or otherwise beneficial (or more likely to be beneficial) in the patient. Stating that the treatment will be effective means that the probability of beneficial therapeutic effect is greater than in a person not having the appropriate presence of the particular kinase activity increasing nucleic acid variance(s) in the kinase domain of the erbB1 gene.

[00150] The treatment will involve the administration of a tyrosine kinase inhibitor. The treatment may involve a combination of treatments, including, but not limited to a tyrosine kinase inhibitor in combination with other tyrosine kinase inhibitors, chemotherapy, radiation, etc..

[00151] Thus, in connection with the administration of a tyrosine kinase inhibitor, a drug which is "effective against" a cancer indicates that administration in a clinically appropriate manner results in a beneficial effect for at least a statistically significant fraction of patients, such as a improvement of symptoms, a cure, a reduction in disease load, reduction in tumor mass or cell numbers, extension of life, improvement in quality of life, or other effect generally recognized as positive by medical doctors familiar with treating the particular type of disease or condition.

[00152] In a preferred embodiment, the compound is an anilinoquinazoline or synthetic anilinoquinazoline. European Patent Publication No. 0566226 discloses anilinoquinazolines which have activity against epidermal growth factor (EGF) receptor tyrosine kinase. It is also known from European Patent Applications Nos. 0520722 and 0566226 that certain 4- anilinoquinazoline derivatives are useful as inhibitors of receptor tyrosine kinases. The very tight structure-activity relationships shown by these compounds suggests a clearly-defined binding mode, where the quinazoline ring binds in the adenine pocket and the anilino ring binds in an adjacent, unique lipophilic pocket. Three 4-anilinoquinazoline analogues (two reversible and one irreversible inhibitor) have been evaluated clinically as anticancer drugs. Denny, Farmaco January-February 2001;56(1-2):51-6. Alternatively, the compound is EKB-569, an inhibitor of EGF receptor kinase (Torrance et al., Nature Medicine, vol. 6, No.

9, Sept. 2000, p. 1024). In a most preferred embodiment, the compound is gefitinib (IRESSA®) or erlotinib (TARCEVA®).

[00153] Treatment targeting cancer cells containing at least one mutant EGFR described herein may be administered alone or in combination with any other appropriate anti-cancer treatment and/or therapeutic agent known to one skilled in the art. In one embodiment, treatment of a pathology, such as a cancer, is provided comprising administering to a subject in need thereof therapeutically effective amounts of a compound that inhibits EGFR kinase activity, such as gefitinib, erlotinib, etc., administered alone or in combination with at least one other anti-cancer agent or therapy. Inhibition of activated protein kinases through the use of targeted small molecule drugs or antibody-based strategies has emerged as an effective approach to cancer therapy. See, e.g., G. D. Demetri et al., N. Engl. J. Med. 347, 472 (2002); B. J. Druker et al., N. Engl. J. Med. 344, 1038 (2001); D. J. Slamon et al., N. Engl. J. Med. 344, 783 (2001).

[00154] In one embodiment, the anti-cancer agent is at least one chemotherapeutic agent. In a related embodiment, the anti-cancer agent is at least one radiotherapy. In a variant embodiment, the anti-cancer therapy is an antiangiogenic therapy (e.g., endostatin, angiostatin, TNP-470, Caplostatin (Stachi-Fainaro et al., Cancer Cell 7(3), 251 (2005)).

[00155] The therapeutic agents may be the same or different, and may be, for example, therapeutic radionuclides, drugs, hormones, hormone antagonists, receptor antagonists, enzymes or proenzymes activated by another agent, autocrines, cytokines or any suitable anti-cancer agent known to those skilled in the art. In one embodiment, the anti-cancer agent is Avastin, an anti-VEGF antibody proven successful in anti-angiogenic therapy of cancer against both solid cancers and hematological malignancies. See, e.g., Ribatti et al. 2003 J Hematother Stem Cell Res. 12(1), 11-22. Toxins also can be used in the methods of the present invention. Other therapeutic agents useful in the present invention include anti-DNA, anti-RNA, radiolabeled oligonucleotides, such as antisense oligonucleotides, anti-protein and anti-chromatin cytotoxic or antimicrobial agents. Other therapeutic agents are known

to those skilled in the art, and the use of such other therapeutic agents in accordance with the present invention is specifically contemplated.

[00156] The antitumor agent may be one of numerous chemotherapy agents such as an alkylating agent, an antimetabolite, a hormonal agent, an antibiotic, an antibody, an anti-cancer biological, gleevec, colchicine, a vinca alkaloid, L-asparaginase, procarbazine, hydroxyurea, mitotane, nitrosoureas or an imidazole carboxamide. Suitable agents are those agents that promote depolarization of tubulin or prohibit tumor cell proliferation. Chemotherapeutic agents contemplated as within the scope of the invention include, but are not limited to, anti-cancer agents listed in the Orange Book of Approved Drug Products With Therapeutic Equivalence Evaluations, as compiled by the Food and Drug Administration and the U.S. Department of Health and Human Services. Nonlimiting examples of chemotherapeutic agents include, e.g., carboplatin and paclitaxel. Treatments targeting EGFR kinasc activity can also be administered together with radiation therapy treatment. Additional anti-cancer treatments known in the art are contemplated as being within the scope of the invention.

[00157] The therapeutic agent may be a chemotherapeutic agent. Chemotherapeutic agents are known in the art and include at least the taxanes, nitrogen mustards, ethylenimine derivatives, alkyl sulfonates, nitrosoureas, triazenes; folic acid analogs, pyrimidine analogs, purine analogs, vinca alkaloids, antibiotics, enzymes, platinum coordination complexes, substituted urea, methyl hydrazine derivatives, adrenocortical suppressants, or antagonists. More specifically, the chemotherapeutic agents may be one or more agents chosen from the non-limiting group of steroids, progestins, estrogens, antiestrogens, or androgens. Even more specifically, the chemotherapy agents may be azaribine, bleomycin, bryostatin-1, busulfan, carmustine, chlorambucil, carboplatin, cisplatin, CPT-11, cyclophosphamide, cytarabine, dacarbazine, dactinomycin, daunorubicin, dexamethasone, diethylstilbestrol, doxorubicin, ethinyl estradiol, etoposide, fluorouracil, fluoxymesterone, gemcitabine, hydroxyprogesterone caproate, hydroxyurea, L-asparaginase, leucovorin, lomustine, mechlorethamine, medroprogesterone acetate, megestrol acetate, melphalan, mercaptopurine,

methotrexate, methotrexate, mithramycin, mitomycin, mitotane, paclitaxel, phenyl butyrate, prednisone, procarbazine, semustine streptozocin, tamoxifen, taxanes, taxol, testosterone propionate, thalidomide, thioguanine, thiotepa, uracil mustard, vinblastine, or vincristine. The use of any combinations of chemotherapy agents is also contemplated. The administration of the chemotherapeutic agent may be before, during or after the administration of a treatment targeting EGFR activity.

[00158] Other suitable therapeutic agents are selected from the group consisting of radioisotope, boron addend, immunomodulator, toxin, photoactive agent or dye, cancer chemotherapeutic drug, antiviral drug, antifungal drug, antibacterial drug, antiprotozoal drug and chemosensitizing agent (See, U.S. Patent Nos. 4,925,648 and 4932,412). Suitable chemotherapeutic agents are described in REMINGTON'S PHARMACEUTICAL SCIENCES, 19th Ed. (Mack Publishing Co. 1995), and in Goodman and Gilman's The Pharmacological Basis of Therapeutics (Goodman et al., Eds. Macmillan Publishing Co., New York, 1980 and 2001 editions). Other suitable chemotherapeutic agents, such as experimental drugs, are known to those of skill in the art. Moreover a suitable therapeutic radioisotope is selected from the group consisting of α-emitters, β-emitters, γ-emitters, Auger electron emitters, neutron capturing agents that emit α-particles and radioisotopes that decay by electron capture. Preferably, the radioisotope is selected from the group consisting of 225Ac, 198Au, 32P, 125I, 131I, 90Y, 186Re, 188Re, 67Cu, 177Lu, 213Bi, 10B, and 211At.

[00159] Where more than one therapeutic agent is used, they may be the same or different. For example, the therapeutic agents may comprise different radionuclides, or a drug and a radionuclide. In a preferred embodiment, treatment targeting EGFR activity inhibits mutant EGFR kinase activity.

[00160] In another embodiment, different isotopes that are effective over different distances as a result of their individual energy emissions are used as first and second therapeutic agents. Such agents can be used to achieve more effective treatment of tumors, and are useful in patients presenting with multiple tumors of differing sizes, as in normal clinical circumstances.

[00161] Few of the available isotopes are useful for treating the very smallest tumor deposits and single cells. In these situations, a drug or toxin may be a

more useful therapeutic agent. Accordingly, in preferred embodiments of the present invention, isotopes are used in combination with non-isotopic species such as drugs, toxins, and neutron capture agents. Many drugs and toxins are known which have cytotoxic effects on cells, and can be used in connection with the present invention. They are to be found in compendia of drugs and toxins, such as the Merck Index, Goodman and Gilman, and the like, and in the references cited above.

[00162] Drugs that interfere with intracellular protein synthesis can also be used in the methods of the present invention; such drugs are known to those skilled in the art and include puromycin, cycloheximide, and ribonuclease.

[00163] The therapeutic methods of the invention may be used for cancer therapy. It is well known that radioisotopes, drugs, and toxins can be conjugated to antibodies or antibody fragments which specifically bind to markers which are produced by or associated with cancer cells, and that such antibody conjugates can be used to target the radioisotopes, drugs or toxins to tumor sites to enhance their therapeutic efficacy and minimize side effects. Examples of these agents and methods are reviewed in Wawrzynczak and Thorpe (in Introduction to the Cellular and Molecular Biology of Cancer, L. M. Franks and N. M. Teich, eds, Chapter 18, pp. 378-410, Oxford University Press. Oxford, 1986), in Immunoconjugates: Antibody Conjugates in Radioimaging and Therapy of Cancer (C. W. Vogel, ed., 3-300, Oxford University Press, N.Y., 1987), in Dillman, R. O. (CRC Critical Reviews in Oncology/Hematology 1:357, CRC Press, Inc., 1984), in Pastan et al. (Cell 47:641, 1986). in Vitetta et al. (Science 238:1098-1104, 1987) and in Brady et al. (Int. J. Rad. Oncol. Biol. Phys. 13:1535-1544, 1987). Other examples of the use of immunoconjugates for cancer and other forms of therapy have been disclosed, inter alia, in U.S. Pat. Nos. 4,331,647, 4,348,376, 4,361,544, 4,468,457, 4,444,744, 4,460,459, 4,460,561 4,624,846, 4,818,709, 4,046,722, 4,671,958, 4,046,784, 5,332,567, 5,443,953, 5,541,297, 5,601,825, 5,635,603, 5,637,288, 5,677,427, 5,686,578, 5,698,178, 5,789,554, 5,922,302, 6,187,287, and 6,319,500.

[00164] Additionally, the treatment methods of the invention can be used in combination with other compounds or techniques for preventing, mitigating or reversing the side effects of certain cytotoxic agents. Examples of such combinations

include, e.g., administration of IL-1 together with an antibody for rapid clearance, as described in e.g., U.S. Pat. No. 4,624,846. Such administration can be performed from 3 to 72 hours after administration of a primary therapeutic treatment targeting EGFR activity in combination with an anti-cancer agent (e.g., with a radioisotope, drug or toxin as the cytotoxic component). This can be used to enhance clearance of the conjugate, drug or toxin from the circulation and to mitigate or reverse myeloid and other hematopoietic toxicity caused by the therapeutic agent.

[00165] In another aspect of the invention, cancer therapy may involve a combination of more than one tumoricidal agent, e.g., a drug and a radioisotope, or a radioisotope and a Boron-10 agent for neutron-activated therapy, or a drug and a biological response modifier, or a fusion molecule conjugate and a biological response modifier. The cytokine can be integrated into such a therapeutic regimen to maximize the efficacy of each component thereof.

[00166] Similarly, certain antileukemic and antilymphoma antibodies conjugated with radioisotopes that are β or α emitters may induce myeloid and other hematopoietic side effects when these agents are not solely directed to the tumor cells. This is observed particularly when the tumor cells are in the circulation and in the blood-forming organs. Concomitant and/or subsequent administration of at least one hematopoietic cytokine (e.g., growth factors, such as colony stimulating factors, such as G-CSF and GM-CSF) is preferred to reduce or ameliorate the hematopoietic side effects, while augmenting the anticancer effects.

[00167] It is well known in the art that various methods of radionuclide therapy can be used for the treatment of cancer and other pathological conditions, as described, e.g., in Harbert, "Nuclear Medicine Therapy", New York, Thieme Medical Publishers, 1087, pp. 1-340. A clinician experienced in these procedures will readily be able to adapt the cytokine adjuvant therapy described herein to such procedures to mitigate any hematopoietic side effects thereof. Similarly, therapy with cytotoxic drugs, administered with treatment targeting EGFR activity, can be used, e.g., for treatment of cancer or other cell proliferative diseases. Such treatment is governed by analogous principles to radioisotope therapy with isotopes or radiolabeled antibodies.

The ordinary skilled clinician will be able to adapt the administration of the additional anti-cancer therapy before, during and/or after the primary anti-cancer therapy.

KITS

[00168] The present invention therefore also provides predictive, diagnostic, and prognostic kits comprising degenerate primers to amplify a target nucleic acid in the kinase domain of the erbB1 gene and instructions comprising amplification protocol and analysis of the results. The kit may alternatively also comprise buffers, enzymes, and containers for performing the amplification and analysis of the amplification products. The kit may also be a component of a screening, diagnostic or prognostic kit comprising other tools such as DNA microarrays. Preferably, the kit also provides one or more control templates, such as nucleic acids isolated from normal tissue sample, and/or a series of samples representing different variances in the kinase domain of the erbB1 gene.

[00169] In one embodiment, the kit provides two or more primer pairs, each pair capable of amplifying a different region of the erbB1 gene (each region a site of potential variance) thereby providing a kit for analysis of expression of several gene variances in a biological sample in one reaction or several parallel reactions.

[00170] Primers in the kits may be labeled, for example fluorescently labeled, to facilitate detection of the amplification products and consequent analysis of the nucleic acid variances.

[00171] In one embodiment, more than one variance can be detected in one analysis. A combination kit will therefore comprise of primers capable of amplifying different segments of the kinase domain of the erbB1 gene. The primers may be differentially labeled, for example using different fluorescent labels, so as to differentiate between the variances.

[00172] The primers contained within the kit may include the following primers: Exon 19 sense primer, 5'- GCAATATCAGCCTTAGGTGCGGCTC-3' (SEQ ID NO: 505); Exon 19 antisense primer, 5'-CATAGAA AGTGAACATTTAGGATGTG-3' (SEQ ID NO: 506); Exon 21 sense primer, 5'-CTAACGTTCG CCAGCCATAAGTCC-3' (SEQ ID NO: 507); and Exon 21

antisense primer, 5'- GCTGCGAGCTCACCCAG AATGTCTGG-3' (SEQ ID NO: 508).

[00173] In a preferred embodiment, the primers are selected from the group consisting of SEQ ID NOS 646-673 (see Tables 5 and 6). These primers have SEQ ID NO 645 on the 5' end of the forward primer and SEQ ID NO 674 on the 5' end of the reverse primers.

Immunodetection Kits

[00174] In further embodiments, the invention provides immunological kits for use in detecting the activation levels of downstream EGFR targets (i.e. STAT3, STAT5, and Akt). Such kits will generally comprise one or more antibodies that have immunospecificity for the phosphorylated form of STAT3, STAT5, or Akt.

[00175] A kit comprising an antibody capable of immunospecifically binding a phosphorylated protein in a mammalian cell selected from the group consisting of phosphorylated Akt, STAT3, and STAT5 proteins and instrctions for using the antibody to examine the mammalian cell for Akt, STAT3 or STAT5 pathway activation is provided in the present invention. In preferred methods, the kit comprises different antibodies, each of which is capable of immunospecifically binding phosphorylated proteins in a mammalian cell selected from the group consisting of phosphorylated Akt, STAT3 or STAT5 proteins.

[00176] The kit generally comprises, a) a pharmaceutically acceptable carrier; b) an antibody directed against phosphorylated STAT3, STAT5, or Akt, in a suitable container means; and c) an immunodetection reagent. Antibodies (monoclonal or polyclonal) are commercially available and may also be prepared by methods known to those of skill in the art, for example, in Current Protocols in Immunology, John Wiley & Sons, Edited by: John E. Coligan, Ada M. Kruisbeek, David H. Margulies, Ethan M. Shevach, Warren Strober, 2001.

[00177] In certain embodiments, the antigen or the antibody may be bound to a solid support, such as a column matrix or well of a microtitre plate. The immunodetection reagents of the kit may take any one of a variety of forms, including those detectable labels that are associated with, or linked to, the given antibody or

antigen itself. Detectable labels that are associated with or attached to a secondary binding ligand are also contemplated. Exemplary secondary ligands are those secondary antibodies that have binding affinity for the first antibody or antigen.

[00178] Suitable assay labels are known in the art and include enzyme labels, such as, glucose oxidase; radioisotopes, such as iodine (^{131}I, ^{125}I, ^{123}I, ^{121}I), carbon (^{14}C), sulfur (^{35}S), tritium (^{3}H), indium (^{115m}In, ^{113m}In, ^{112}In, ^{111}In), and technetium (^{99}Tc, ^{99m}Tc), thallium (^{201}Ti), gallium (^{68}Ga, ^{67}Ga), palladium (^{103}Pd), molybdenum (^{99}Mo), xenon (^{133}Xe), fluorine (^{18}F), ^{153}Sm, ^{177}Lu, ^{159}Gd, ^{149}Pm, ^{140}La, ^{175}Yb, ^{166}Ho, ^{90}Y, ^{47}Sc, ^{186}Re, ^{188}Re, ^{142}Pr, ^{105}Rh, ^{97}Ru; luminescent labels, such as luminol; and fluorescent labels, such as fluorescein and rhodamine, and biotin.

[00179] Further suitable immunodetection reagents for use in the present kits include the two-component reagent that comprises a secondary antibody that has binding affinity for the first antibody or antigen, along with a third antibody that has binding affinity for the second antibody, wherein the third antibody is linked to a detectable label.

[00180] A number of exemplary labels are known in the art and all such labels may be employed in connection with the present invention. Radiolabels, nuclear magnetic spin-resonance isotopes, fluorescent labels and enzyme tags capable of generating a colored product upon contact with an appropriate substrate are suitable examples.

[00181] The kits may contain antibody-label conjugates either in fully conjugated form, in the form of intermediates, or as separate moieties to be conjugated by the user of the kit.

[00182] The kits may further comprise a suitably aliquoted composition of an antigen whether labeled or unlabeled, as may be used to prepare a standard curve for a detection assay or as a positive control.

[00183] The kits of the invention, regardless of type, will generally comprise one or more containers into which the biological agents are placed and, preferably, suitable aliquoted. The components of the kits may be packaged either in aqueous media or in lyophilized form.

[00184] The immunodetection kits of the invention may additionally contain one or more of a variety of other cancer marker antibodies or antigens, if so desired. Such kits could thus provide a panel of cancer markers, as may be better used in testing a variety of patients. By way of example, such additional markers could include, other tumor markers such as PSA, SeLe (X), HCG, as well as p53, cyclin D1, p16, tyrosinase, MAGE, BAGE, PAGE, MUC18, CEA, p27, [bgr]HCG or other markers known to those of skill in the art.

[00185] The container means of the kits will generally include at least one vial, test tube, flask, bottle, or even syringe or other container means, into which the antibody or antigen may be placed, and preferably, suitably aliquoted. Where a second or third binding ligand or additional component is provided, the kit will also generally contain a second, third or other additional container into which this ligand or component may be placed.

[00186] The kits of the present invention will also typically include a means for containing the antibody, antigen, and any other reagent containers in close confinement for commercial sale. Such containers may include injection or blow-molded plastic containers into which the desired vials are retained.

[00187] The methods of the present invention also encompass the identification of compounds that interfere with the kinase activity of a variant form of the EGFR. The variant EGFR comprises at least one variance in its kinase domain. Such compounds may, for example, be tyrosine kinase inhibitors. Methods for identifying compounds that interfere with the kinase activity of a receptor are generally known to those of skill in the art and are further described in, for example, for example, Dhanabal et al., Cancer Res. 59:189-197 (1999); Xin et al., J. Biol. Chem. 274:9116-9121 (1999); Sheu et al., Anticancer Res. 18:4435-4441; Ausprunk et al., Dev. Biol. 38:237-248 (1974); Gimbrone et al., J. Natl. Cancer Inst. 52:413-427; Nicosia et al., *In vitro* 18:538-549, incorporated herein by reference. In general, compounds are identified, using the methods disclosed herein, that interfere with the enhanced kinase activity characteristic of at least one variance in the kinase domain of the erbB1 gene.

Solid Support

[00188] In another embodiment, the invention provides a kit for practicing the methods of the invention. In one embodiment, a kit for the detection of variances in the kinase domain of erbB1 gene on a solid support is described. The kit can include, e.g. the materials and reagents for detecting a plurality of variances in one assay. The kit can include e.g. a solid support, oligonucleotide primers for a specific set of target polynucleotides, polymerase chain reaction reagents and components, e.g. enzymes for DNA synthesis, labeling materials, and other buffers and reagents for washing. The kit may also include instructions for use of the kit to amplify specific targets on a solid support. Where the kit contains a prepared solid support having a set of primers already fixed on the solid support, e.g. for amplifying a particular set of target polynucleotides, the design and construction of such a prepared solid support is described above. The kit also includes reagents necessary for conducting a PCR on a solid support, for example using an in situ-type or solid phase type PCR procedure where the support is capable of PCR amplification using an in situ-type PCR machine. The PCR reagents, included in the kit, include the usual PCR buffers, a thermostable polymerase (e.g. Taq DNA polymerase), nucleotides (e.g. dNTPs), and other components and labeling molecules (e.g. for direct or indirect labeling as described above). The kits can be assembled to support practice of the PCR amplification method using immobilized primers alone or, alternatively, together with solution phase primers.

[00189] Alternatively, the kit may include a solid support with affixed oligonucleotides specific to any number of EGFR variances, further defined in Figures 4A-4C and Figures 7 and 8. A test biological sample may be applied to the solid support, under selective hybridization conditions, for the determination of the presence or absence of variances in the kinase domain of erbB1.

[00190] The methods of the present invention also encompass the identification of compounds that interfere with the kinase activity of a variant form of the EGFR. The variant EGFR comprises at least one variance in its kinase domain. However, in an alternative embodiment, the variant EGFR comprises a secondary

mutation that confers resistance to a first TKI e.g., gefitinib or erlotinib. Such compounds may, for example, be tyrosine kinase inhibitors. Methods for identifying compounds that interfere with the kinase activity of a receptor are generally known to those of skill in the art and are further described in, for example, for example, Dhanabal et al., Cancer Res. 59:189-197 (1999); Xin et al., J. Biol. Chem. 274:9116-9121 (1999); Sheu et al., Anticancer Res. 18:4435-4441; Ausprunk et al., Dev. Biol. 38:237-248 (1974); Gimbrone et al., J. Natl. Cancer Inst. 52:413-427; Nicosia et al., *In vitro* 18:538-549, incorporated herein by reference. In general, compounds are identified, using the methods disclosed herein, that interfere with the enhanced kinase activity characteristic of at least one variance in the kinase domain of the erbB1 gene. Such known variances are described in Figures 4, 7, 8 and Table 2.

[00191] Once identified, such compounds are administered to patients in need of EGFR targeted treatment, for example, patients affected with or at risk for developing cancer.

[00192] The route of administration may be intravenous (I.V.), intramuscular (I.M.), subcutaneous (S.C.), intradermal (I.D.), intraperitoneal (I.P.), intrathecal (I.T.), intrapleural, intrauterine, rectal, vaginal, topical, intratumor and the like. The compounds of the invention can be administered parenterally by injection or by gradual infusion over time and can be delivered by peristaltic means.

[00193] Administration may 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 bile salts and fusidic acid derivatives. In addition, detergents may be used to facilitate permeation. Transmucosal administration may be through nasal sprays, for example, or using suppositories. For oral administration, the compounds of the invention are formulated into conventional oral administration forms such as capsules, tablets and tonics.

[00194] For topical administration, the pharmaceutical composition (inhibitor of kinase activity) is formulated into ointments, salves, gels, or creams, as is generally known in the art.

[00195] The therapeutic compositions of this invention are conventionally administered intravenously, as by injection of a unit dose, for example. The term "unit dose" when used in reference to a therapeutic composition of the present invention refers to physically discrete units suitable as unitary dosage for the subject, each unit containing a predetermined quantity of active material calculated to produce the desired therapeutic effect in association with the required diluent; i.e., carrier, or vehicle.

[00196] The compositions are administered in a manner compatible with the dosage formulation, and in a therapeutically effective amount. The quantity to be administered and timing depends on the subject to be treated, capacity of the subject's system to utilize the active ingredient, and degree of therapeutic effect desired. Precise amounts of active ingredient required to be administered depend on the judgment of the practitioner and are peculiar to each individual.

[00197] The tyrosine kinase inhibitors useful for practicing the methods of the present invention are described herein. Any formulation or drug delivery system containing the active ingredients, which is suitable for the intended use, as are generally known to those of skill in the art, can be used. Suitable pharmaceutically acceptable carriers for oral, rectal, topical or parenteral (including inhaled, subcutaneous, intraperitoneal, intramuscular and intravenous) administration are known to those of skill in the art. The carrier must be pharmaceutically acceptable in the sense of being compatible with the other ingredients of the formulation and not deleterious to the recipient thereof.

[00198] As used herein, the terms "pharmaceutically acceptable", "physiologically tolerable" and grammatical variations thereof, as they refer to compositions, carriers, diluents and reagents, are used interchangeably and represent that the materials are capable of administration to or upon a mammal without the production of undesirable physiological effects.

[00199] Formulations suitable for parenteral administration conveniently include sterile aqueous preparation of the active compound which is preferably isotonic with the blood of the recipient. Thus, such formulations may conveniently contain distilled water, 5% dextrose in distilled water or saline. Useful formulations

also include concentrated solutions or solids containing the compound which upon dilution with an appropriate solvent give a solution suitable for parental administration above.

[00200] For enteral administration, a compound can be incorporated into an inert carrier in discrete units such as capsules, cachets, tablets or lozenges, each containing a predetermined amount of the active compound; as a powder or granules; or a suspension or solution in an aqueous liquid or non-aqueous liquid, e.g., a syrup, an elixir, an emulsion or a draught. Suitable carriers may be starches or sugars and include lubricants, flavorings, binders, and other materials of the same nature.

[00201] A tablet may be made by compression or molding, optionally with one or more accessory ingredients. Compressed tablets may be prepared by compressing in a suitable machine the active compound in a free-flowing form, e.g., a powder or granules, optionally mixed with accessory ingredients, e.g., binders, lubricants, inert diluents, surface active or dispersing agents. Molded tablets may be made by molding in a suitable machine, a mixture of the powdered active compound with any suitable carrier.

[00202] A syrup or suspension may be made by adding the active compound to a concentrated, aqueous solution of a sugar, e.g., sucrose, to which may also be added any accessory ingredients. Such accessory ingredients may include flavoring, an agent to retard crystallization of the sugar or an agent to increase the solubility of any other ingredient, e.g., as a polyhydric alcohol, for example, glycerol or sorbitol.

[00203] Formulations for rectal administration may be presented as a suppository with a conventional carrier, e.g., cocoa butter or Witepsol S55 (trademark of Dynamite Nobel Chemical, Germany), for a suppository base.

[00204] Formulations for oral administration may be presented with an enhancer. Orally-acceptable absorption enhancers include surfactants such as sodium lauryl sulfate, palmitoyl carnitine, Laureth-9, phosphatidylcholine, cyclodextrin and derivatives thereof; bile salts such as sodium deoxycholate, sodium taurocholate, sodium glycochlate, and sodium fusidate; chelating agents including EDTA, citric acid and salicylates; and fatty acids (e.g., oleic acid, lauric acid, acylcarnitines, mono- and diglycerides). Other oral absorption enhancers include benzalkonium chloride,

benzethonium chloride, CHAPS (3-(3-cholamidopropyl)-dimethylammonio-1-propanesulfonate), Big-CHAPS (N, N-bis(3-D-gluconamidopropyl)-cholamide), chlorobutanol, octoxynol-9, benzyl alcohol, phenols, cresols, and alkyl alcohols. An especially preferred oral absorption enhancer for the present invention is sodium lauryl sulfate.

[00205] Alternatively, the compound may be administered in liposomes or microspheres (or microparticles). Methods for preparing liposomes and microspheres for administration to a patient are well known to those of skill in the art. U.S. Pat. No. 4,789,734, the contents of which are hereby incorporated by reference, describes methods for encapsulating biological materials in liposomes. Essentially, the material is dissolved in an aqueous solution, the appropriate phospholipids and lipids added, along with surfactants if required, and the material dialyzed or sonicated, as necessary. A review of known methods is provided by G. Gregoriadis, Chapter 14, "Liposomes," Drug Carriers in Biology and Medicine, pp. 287-341 (Academic Press, 1979).

[00206] Microspheres formed of polymers or proteins are well known to those skilled in the art, and can be tailored for passage through the gastrointestinal tract directly into the blood stream. Alternatively, the compound can be incorporated and the microspheres, or composite of microspheres, implanted for slow release over a period of time ranging from days to months. See, for example, U.S. Pat. Nos. 4,906,474, 4,925,673 and 3,625,214, and Jein, TIPS 19:155-157 (1998), the contents of which are hereby incorporated by reference.

[00207] In one embodiment, the tyrosine kinase inhibitor of the present invention can be formulated into a liposome or microparticle which is suitably sized to lodge in capillary beds following intravenous administration. When the liposome or microparticle is lodged in the capillary beds surrounding ischemic tissue, the agents can be administered locally to the site at which they can be most effective. Suitable liposomes for targeting ischemic tissue are generally less than about 200 nanometers and are also typically unilamellar vesicles, as disclosed, for example, in U.S. Pat. No. 5,593,688 to Baldeschweiler, entitled "Liposomal targeting of ischemic tissue," the contents of which are hereby incorporated by reference.

[00208] Preferred microparticles are those prepared from biodegradable polymers, such as polyglycolide, polylactide and copolymers thereof. Those of skill in the art can readily determine an appropriate carrier system depending on various factors, including the desired rate of drug release and the desired dosage.

[00209] In one embodiment, the formulations are administered via catheter directly to the inside of blood vessels. The administration can occur, for example, through holes in the catheter. In those embodiments wherein the active compounds have a relatively long half life (on the order of 1 day to a week or more), the formulations can be included in biodegradable polymeric hydrogels, such as those disclosed in U.S. Pat. No. 5,410,016 to Hubbell et al. These polymeric hydrogels can be delivered to the inside of a tissue lumen and the active compounds released over time as the polymer degrades. If desirable, the polymeric hydrogels can have microparticles or liposomes which include the active compound dispersed therein, providing another mechanism for the controlled release of the active compounds.

[00210] The formulations may conveniently be presented in unit dosage form and may be prepared by any of the methods well known in the art of pharmacy. All methods include the step of bringing the active compound into association with a carrier which constitutes one or more accessory ingredients. In general, the formulations are prepared by uniformly and intimately bringing the active compound into association with a liquid carrier or a finely divided solid carrier and then, if necessary, shaping the product into desired unit dosage form.

[00211] The formulations may further include one or more optional accessory ingredient(s) utilized in the art of pharmaceutical formulations, e.g., diluents, buffers, flavoring agents, binders, surface active agents, thickeners, lubricants, suspending agents, preservatives (including antioxidants) and the like.

[00212] Compounds of the present methods may be presented for administration to the respiratory tract as a snuff or an aerosol or solution for a nebulizer, or as a microfine powder for insufflation, alone or in combination with an inert carrier such as lactose. In such a case the particles of active compound suitably have diameters of less than 50 microns, preferably less than 10 microns, more preferably between 2 and 5 microns.

[00213] Generally for nasal administration a mildly acid pH will be preferred. Preferably the compositions of the invention have a pH of from about 3 to 5, more preferably from about 3.5 to about 3.9 and most preferably 3.7. Adjustment of the pH is achieved by addition of an appropriate acid, such as hydrochloric acid.

[00214] The preparation of a pharmacological composition that contains active ingredients dissolved or dispersed therein is well understood in the art and need not be limited based on formulation. Typically such compositions are prepared as injectables either as liquid solutions or suspensions, however, solid forms suitable for solution, or suspensions, in liquid prior to use can also be prepared. The preparation can also be emulsified.

[00215] The active ingredient can be mixed with excipients which are pharmaceutically acceptable and compatible with the active ingredient and in amounts suitable for use in the therapeutic methods described herein. Suitable excipients are, for example, water, saline, dextrose, glycerol, ethanol or the like and combinations thereof. In addition, if desired, the composition can contain minor amounts of auxiliary substances such as wetting or emulsifying agents, pH buffering agents and the like which enhance the effectiveness of the active ingredient.

[00216] The kinase inhibitor of the present invention can include pharmaceutically acceptable salts of the components therein. Pharmaceutically acceptable salts include the acid addition salts (formed with the free amino groups of the polypeptide) that are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, tartaric, mandelic and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine and the like.

[00217] Physiologically tolerable carriers are well known in the art. Exemplary of liquid carriers are sterile aqueous solutions that contain no materials in addition to the active ingredients and water, or contain a buffer such as sodium phosphate at physiological pH value, physiological saline or both, such as phosphate-buffered saline. Still further, aqueous carriers can contain more than one buffer salt, as

well as salts such as sodium and potassium chlorides, dextrose, polyethylene glycol and other solutes.

[00218] Liquid compositions can also contain liquid phases in addition to and to the exclusion of water. Exemplary of such additional liquid phases are glycerin, vegetable oils such as cottonseed oil, and water-oil emulsions.

Predicting Mutations

[00219] In another embodiment, the present invention discloses a method to predict variances in the erbB1 gene following treatment with a tyrosine kinase inhibitor. It is generally known that response to cancer treatment with a tyrosine kinase inhibitor is often followed by resistance to that or other similar compounds. Such resistance is thought to arise through the acquisition of mutations in the drug target, for example in the EGFR. The ability to predict (and select) such mutations will allow for better treatment options and fewer relapses.

[00220] In one embodiment of the present invention, DNA encoding the EGFR kinase domain is isolated and sequenced from a tumor sample of cancer patients that have responded to gefitinib (or a similar EGFR targeting treatment) but have subsequently relapsed. The relapse in such patients is expected to involve the acquisition of secondary mutations within the EGFR kinase domain. Compounds that target, and inhibit the kinase activity of, these newly defined mutations are then identified using methods disclosed herein. Such compounds may be used alone, or in combination with other known EGFR targeting treatments, to treat cancer patients with primary or secondary (as above) mutations in the kinase domain of EGFR.

[00221] In one embodiment, predicting variances in the kinase (catalytic) domain of the EGFR (erbB1 gene) is done *in vitro*. In this method, cells, e.g. fibroblast cells, are stably transfected with cDNAs containing kinase domain mutations that have been identified in human cancer cell lines. For example, the cells may be transfected with an EGFR that bears a mutation such as SEQ ID NO:495, further described in Figure 4A, or with any number of identified or as yet unidentified kinase domain-mutated EGFRs. The transfection of kinase domain-mutated EGFRs into cells will result in aberrant proliferation of the cells in culture. Methods of stable

transfection are known to those of skill in the art and are further defined in Current Protocols in Molecular Biology by F. M. Ausubel, R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, K. Struhl and V. B. Chanda (Editors), John Wiley & Sons., 2004, incorporated herein by reference. The transfected cells are then given an effective, yet sub-lethal, dose of a drug, preferably a tyrosine kinase inhibitor, predicted to inhibit cellular proliferation. In a preferred embodiment, the drug is an anilinoquinazoline, synthetic anilinoquinazoline, gefitinib or erlotinib. The cells are serially passaged in the presence of drug and subclones that survive are selected. Over many generations, cells that survive (i.e. are resistant to the compound), are selected and analyzed for variances in the erbB1 gene. Secondary variances can thus be predicted to occur following repeated treatment with a tyrosine kinase inhibitor *in vivo*.

[00222] Alternatively, cells are transfected with gefitinib-resistant mutant cDNA derived from human NSCLC cell lines, for example, NCI-1650 and NCI-1975. Each cell line has a heterozygous mutation with the kinase domain of EGFR, and is, therefore, expected to be sensitive to gefitinib. The EGFR mutation in NCI-1650 consists of an in-frame deletion of 15 nucleotides at position 2235-2249 (delLE746-A750) within exon 19, while NCI-1975 has a missense mutation within exon 21 that substitutes a G for T at nucleotide 2573 (L858R). As shown herein, the L858R mutation in NCI-H1975 is activating and confers increased sensitivity to gefitinib in vitro. Other cancer cell lines that harbor EGFR kinase domain mutations may be utilized. The cancer cell lines may include lung cancer as well as other cancers that are found to harbor such mutations.

[00223] The cells may be treated with a mutagen in order to increase the frequency with which cells acquire secondary mutations. A mutagen may induce mutations at different frequencies depending upon the dosage regimen, mode of delivery, and the developmental stage of the organism or cell upon mutagen administration, all parameters of which are disclosed in the prior art for different mutagens or mutagenesis techniques. The mutagen may be an alkylating agent, such as ethyl methanesulfonate (EMS), N-ethyl-N-nitrosourea (ENU) or N-methyl-N-nitrosourea (MNU). Alternatively, the mutagen may be, for example, phocarbaxine

hydrochloride (Prc), methyl methanesulfonate (MeMS), chlorambucil (Chl), melphalan, porcarbazine hydrochloride, cyclophosphamide (Cp), diethyl sulfate (Et_2SO_4), acrylamide monomer (AA), triethylene melamin (TEM), nitrogen mustard, vincristine, dimethylnitrosamine, N-methyl-N'-nitro-Nitrosoguanidine (MNNG), 7,12 dimethylbenz(a)anthracene (DMBA), ethylene oxide, hexamethylphosphoramide, bisulfan, and ethyl methanesulforate (EtMs). Methods of treating cells with mutagens is described, for example, in U.S. 6,015,670, incorporated herein by reference. Following mutagenesis, cells (i.e. transfected with variant EGFR or human cancer cell line derived) can be cultured in gefitinib-supplemented medium to select for the outgrowth of resistant clones. Subcultivation of individual clones can be followed, for example, by nucleotide sequence determination of the EGFR gene following specific PCR-mediated amplification of genomic DNA corresponding to the EGFR kinase domain.

[00224] In another embodiment, cells (with an EGFR variance) are serially passaged in the presence of gradually increasing concentrations of gefitinib (or a similar tyrosine kinase inhibitor) over a course of several weeks or months in order to select for the spontaneous acquisition of mutations within the EGFR gene that confer resistance to gefitinib. Selected cells (that continue to proliferate at relatively high gefitinib concentration) can be isolated as colonies, and mutations will be identified as described above. Such variances can thus be predicted to occur following repeated treatment with a tyrosine kinase inhibitor *in vivo*. See, for example, Scappini et al., Cancer, April 1, 2004, Vol. 100, pg. 1459, incorporated herein by reference.

[00225] In yet another embodiment, a variant form of the EGFR gene can be propagated in a DNA repair-deficent bacterial strain before re-introducing it into stably selected cell lines. Replication in such bacteria will enhance the frequency of mutagenesis. Alternatively, "error-prone" PCR can be utilized to enhance the frequency of mutations in the cloned EGFR DNA in vitro, using standard methods, known to those of skill in the art.

[00226] In another embodiment, predicting variances in the kinase domain of the erbB1 gene is done *in vivo*. For example, a kinase activity increasing variant form of the erbB1 gene is transfected into an animal, i.e. a mouse, generating a

cancer model. The animal is then treated with an effective dose of a compound, preferably an anilinoquinazoline, synthetic anilinoquinazoline, gefitinib or erlotinib. Upon repeated exposure to the compound, the cancer is initially inhibited. As in humans treated with such compounds, tumor cells in the animal acquire mutations which make them resistant to such treatment. The methods of the present invention allow for the isolation and characterization of the erbB 1 gene in such resistant tumors. Compounds that specifically target these newly characterized variances are useful in the treatment of patients suspected of carrying such a mutated erbB1 gene. Such patients include, for example, patients who initially respond to therapy with a tyrosine kinase inhibitor, but subsequently fail to respond to the same or similar compound.

[00227] Methods of creating an animal model are known to those of skill in the art and are further defined in e.g., Ohashi et al., Cell, 65:305-317 (1991); Adams et al., Nature, 325:223-228 (1987); and Roman et al., Cell, 61:383-396 (1990)., incorporated herein by reference. In the case of fertilized oocytes, the preferred method of transgene introduction is by microinjection, see, e.g., Leder et al., U.S. Pat. Nos. 4,736,866 and 5,175,383, which are incorporated herein by reference, whereas in the case of embryonic stem (ES) cells, the preferred method is electroporation. However, other methods including viral delivery systems such as retroviral infection, or liposomal fusion can be used. The isolation and characterization of nucleic acid is described above and in the examples.

[00228] The above-identified kinase activity increasing variances in the erbB1 gene may be screened for in patients (diagnostically or prognostically), using the methods of the present invention. The presence or absence of such mutations may then be used as a criteria for determining ones sensitivity to treatment with an EGFR targeting compound, such as, for example, a tyrosine kinase inhibitor.

[00229] Compounds that specifically target these newly defined variances, whether detected in vivo or in vitro, can be selected using techniques known in the art and discussed herein. Candidate drug screening assays may be used to identify bioactive candidate agents that inhibit the activity of variant forms of EGFR. Of particular interest are screening assays for agents that have a low toxicity for human cells. A wide variety of assays may be used for this purpose, including labeled in vitro

protein-protein binding assays, electrophoretic mobility shift assays, enzyme activity assays, immunoassays for protein binding, and the like. The purified mutant EGFR protein may also be used for determination of three-dimensional crystal structure, which can be used for modeling intermolecular interactions, transporter function, etc. Such compounds may be, for example, tyrosine kinase inhibitors, antibodies, aptamers, siRNAs, and vectors that inhibit the kinase activity of EGFR.

[00230] In another embodiment, compounds useful in the method of the present invention are antibodies which interfere with kinase signaling via the mutant EGFR, including monoclonal, chimeric humanized, and recombinant antibodies and fragment thereof which are characterized by their ability to inhibit the kinase activity of the EGFR and which have low toxicity.

[00231] Neutralizing antibodies are readily raised in animals such as rabbits or mice by immunization with an EGFR with at least one nucleic acid variance in its kinase domain. Immunized mice are particularly useful for providing sources of B cells for the manufacture of hybridomas, which in turn are cultured to produce large quantities of anti-EGFR monoclonal antibodies. Chimeric antibodies are immunoglobin molecules characterized by two or more segments or portions derived from different animal species. Generally, the variable region of the chimeric antibody is derived from a non-human mammalian antibody, such as murine monoclonal antibody, and the immunoglobin constant region is derived from a human immunoglobin molecule. Preferably, both regions and the combination have low immunogenicity as routinely determined. Humanized antibodies are immunoglobin molecules created by genetic engineering techniques in which the murine constant regions are replaced with human counterparts while retaining the murine antigen binding regions. The resulting mouse-human chimeric antibody should have reduced immunogenicity and improved pharmacokinetics in humans. Preferred examples of high affinity monoclonal antibodies and chimeric derivatives thereof, useful in the methods of the present invention, are described in the European Patent Application EP 186,833; PCT Patent Application WO 92/16553; and US Patent No. 6,090,923.

[00232] Existing or newly identified compounds as described above are useful in the treatment of patients carrying primary and/or secondary EGFR mutations.

[00233] In a preferred embodiment, the compound is an inhibitor of the tyrosine kinase activity of an EGFR with at least one variance in its kinase domain, particularly small molecule inhibitors having selective action on "mutated" EGFRs as compared to other tyrosine kinases. Inhibitors of EGFR include, but are not limited to, tyrosine kinase inhibitors such as quinazolines, such as PID 153035, 4-(3-chloroanilino) quinazoline, or CP- 358,774, pyridopyrimidines, pyrimidopyrimidines, pyrrolopyrimidines, such as CGP 59326, CGP 60261 and CGP 62706, and pyrazolopyrimidines, 4- (phenylamino)-7H- pyrrolo[2,3-d] pyrimidines (Traxler et al., (1996) J. Med Chem 39:2285-2292), curcumin (diferuloyl methane) (Laxmin arayana, et al., (1995), Carcinogen 16:1741-1745), 4,5-bis (4- fluoroanilino) phthalimide (Buchdunger et al. (1995) Clin. Cancer Res. 1:813-821; Dinney et al. (1997) Clin. Cancer Res. 3:161-168); tyrphostins containing nitrothiophene moieties (Brunton et al. (1996) Anti Cancer Drug Design 11:265-295); the protein kinase inhibitor ZD-1 839 (AstraZeneca); CP-358774 (Pfizer, Inc.); PD-01 83805 (Warner-Lambert), EKB-569 (Torrance et al., Nature Medicine, Vol. 6, No. 9, Sept. 2000, p. 1024), HKI-272 and HKI-357 (Wyeth); or as described in International patent application W099/09016 (American Cyanamid); W098/43960 (American Cyanamid); W097/38983 (Warener Labert); W099/06378 (Warner Lambert); W099/06396 (Warner Lambert) ; W096/30347 (Pfizer, Inc.); W096/33978 (Zeneca); W096/33977 (Zeneca); and W096/33980) Zeneca; all herein incorporated by reference.

[00234] In another embodiment, an antisense strategy may be used to interfere with the kinase activity of a variant EGFR. This approach may, for instance, utilize antisense nucleic acids or ribozymes that block translation of a specific mRNA, either by masking that mRNA with an antisense nucleic acid or cleaving it with a ribozyme. For a general discussion of antisense technology, see, e.g., Antisense DNA and RNA, (Cold Spring Harbor Laboratory, D. Melton, ed., 1988).

[00235] Reversible short inhibition of variant EGFR gene transcription may also be useful. Such inhibition can be achieved by use of siRNAs. RNA interference

(RNAi) technology prevents the expression of genes by using small RNA molecules such as small interfering RNAs (siRNAs). This technology in turn takes advantage of the fact that RNAi is a natural biological mechanism for silencing genes in most cells of many living organisms, from plants to insects to mammals (McManus et al., Nature Reviews Genetics, 2002, 3(10) p. 737). RNAi prevents a gene from producing a functional protein by ensuring that the molecule intermediate, the messenger RNA copy of the gene is destroyed. siRNAs can be used in a naked form and incorporated in a vector, as described below. One can further make use of aptamers to specifically inhibit variant EGFR gene transcription, see, for example, U.S. Patent 6,699,843. Aptamers useful in the present invention may be identified using the SELEX process. The methods of SELEX have been described in, for example, U.S. Patent Nos. 5,707,796, 5,763,177, 6,011,577, 5,580,737, 5,567,588, and 5,660,985.

[00236] An "antisense nucleic acid" or "antisense oligonucleotide" is a single stranded nucleic acid molecule, which, on hybridizing under cytoplasmic conditions with complementary bases in a RNA or DNA molecule, inhibits the latter's role. If the RNA is a messenger RNA transcript, the antisense nucleic acid is a countertranscript or mRNA-interfering complementary nucleic acid. As presently used, "antisense" broadly includes RNA-RNA interactions, RNA- DNA interactions, ribozymes, RNAi, aptamers and Rnase-H mediated arrest.

[00237] Ribozymes are RNA molecules possessing the ability to specifically cleave other single stranded RNA molecules in a manner somewhat analogous to DNA restriction endonucleases. Ribozymes were discovered from the observation that certain mRNAs have the ability to excise their own introns. By modifying the nucleotide sequence of these ribozymes, researchers have been able to engineer molecules that recognize specific nucleotide sequences in an RNA molecule and cleave it (Cech, 1989, Science 245(4915) p. 276). Because they are sequence-specific, only mRNAs with particular sequences are inactivated.

[00238] Antisense nucleic acid molecules can be encoded by a recombinant gene for expression in a cell (e.g., U.S. patent No 5,814,500; U.S. 5,811, 234), or alternatively they can be prepared synthetically (e.g., U.S. patent No 5,780,607).

[00239] The present invention further provides methods of treating patients with cancer. In particular, patients with at least one nucleic acid variance in the kinase domain of EGFR. The treatment method comprises administering an siRNA-containing composition to a patient within an appropriate time window. The siRNAs may be chemically synthesized, produced using in vitro transcription, etc. In addition, the siRNA molecule can be customized to individual patients in such a way as to correspond precisely to the mutation identified in their tumor. Since siRNA can discriminate between nucleotide sequences that differ by only a single nucleotide, it is possible to design siRNAs that uniquely target a mutant form of the EGFR gene that is associated with either a single nucleotide substitution or a small deletion of several nucleotides—both of which have been identified in tumors as described herein. SiRNAs have been described in Brummelkamp et al., Science 296; 550-553, 2002, Jaque et al., Nature 418; 435-438, 2002, Elbashir S. M. et al. (2001) Nature, 411: 494-498, McCaffrey et al. (2002), Nature, 418: 38-39; Xia H. et al. (2002), Nat. Biotech. 20: 1006-1010, Novina et al. (2002), Nat. Med. 8: 681-686, and U.S. Application No. 20030198627.

[00240] An important advantage of such a therapeutic strategy relative to the use of drugs such as gefitinib, which inhibit both the mutated receptor and the normal receptor, is that siRNA directed specifically against the mutated EGFR should not inhibit the wildtype EGFR. This is significant because it is generally believed that the "side effects" of gefitinib treatment, which include diarrhea and dermatitis, are a consequence of inhibition of EGFR in normal tissues that require its function.

[00241] The delivery of siRNA to tumors can potentially be achieved via any of several gene delivery "vehicles" that are currently available. These include viral vectors, such as adenovirus, lentivirus, herpes simplex virus, vaccinia virus, and retrovirus, as well as chemical-mediated gene delivery systems (for example, liposomes), or mechanical DNA delivery systems (DNA guns). The oligonucleotides to be expressed for such siRNA-mediated inhibition of gene expression would be between 18 and 28 nucleotides in length.

[00242] In another embodiment, the compounds are antisense molecules specific for human sequences coding for an EGFR having at least one variance in its

kinase domain. The administered therapeutic agent may be an antisense oligonucleotides, particularly synthetic oligonucleotides; having chemical modifications from native nucleic acids, or nucleic acid constructs that express such anti-sense molecules as RNA. The antisense sequence is complementary to the mRNA of the targeted EGFR genes, and inhibits expression of the targeted gene products (see e.g. Nyce et al. (1997) Nature 385:720). Antisense molecules inhibit gene expression by reducing the amount of mRNA available for translation, through activation of RNAse H or steric hindrance. One or a combination of antisense molecules may be administered, where a combination may comprise multiple different sequences from a single targeted gene, or sequences that complement several different genes.

[00243] A preferred target gene is an EGFR with at least one nucleic acid variance in its kinase domain. The gene sequence is incorporated herein, such as, for example, in Figure 5. Generally, the antisense sequence will have the same species of origin as the animal host.

[00244] Antisense molecules may be produced by expression of all or a part of the target gene sequence in an appropriate vector, where the vector is introduced and expressed in the targeted cells. The transcriptional initiation will be oriented such that the antisense strand is produced as an RNA molecule.

[00245] The anti-sense RNA hybridizes with the endogenous sense strand mRNA, thereby blocking expression of the targeted gene. The native transcriptional initiation region, or an exogenous transcriptional initiation region may be employed. The promoter may be introduced by recombinant methods in vitro, or as the result of homologous integration of the sequence into a chromosome. Many strong promoters that are active in muscle cells are known in the art, including the O-actin promoter, SV40 early and late promoters, human cytomegalovirus promoter, retroviral LTRs, etc. Transcription vectors generally have convenient restriction sites located near the promoter sequence to provide for the insertion of nucleic acid sequences. Transcription cassettes maybe prepared comprising a transcription initiation region, the target gene or fragment thereof, and a transcriptional termination region. The transcription cassettes may be introduced into a variety of vectors, e.g. plasmid;

retrovirus, e.g. lentivirus; adenovirus; and the like, where the vectors are able to transiently or stably be maintained in cells, usually for a period of at least about one day, more usually for a period of at least about several days.

[00246] Aptamers are also useful. Aptamers are a promising new class of therapeutic oligonucleotides or peptides and are selected in vitro to specifically bind to a given target with high affinity, such as for example ligand receptors. Their binding characteristics are likely a reflection of the ability of oligonucleotides to form three dimensional structures held together by intramolecular nucleobase pairing. Aptamers are synthetic DNA, RNA or peptide sequences which may be normal and modified (e.g. peptide nucleic acid (PNA), thiophophorylated DNA, etc) that interact with a target protein, ligand (lipid, carbohydrate, metabolite, etc). In a further embodiment, RNA aptamers specific for a variant EGFR can be introduced into or expressed in a cell as a therapeutic.

[00247] Peptide nucleic acids (PNAs) are compounds that in certain respects are similar to oligonucleotides and their analogs and thus may mimic DNA and RNA. In PNA, the deoxyribose backbone of oligonucleotides has been replaced by a pseudo-peptide backbone (Nielsen et al. 1991 Science 254, 1457-1500). Each subunit, or monomer, has a naturally occurring or non-naturally occurring nucleobase attached to this backbone. One such backbone is constructed of repeating units of N-(2-aminoethyl) glycine linked through amide bonds. PNA hybridises with complementary nucleic acids through Watson and Crick base pairing and helix formation. The Pseudo-peptide backbone provides superior hybridization properties (Egholm et al. Nature (1993) 365, 566-568), resistance to enzymatic degradation (Demidov et al. Biochem. Pharmacol. (1994) 48, 1310-1313) and access to a variety of chemical modifications (Nielsen and Haaima Chemical Society Reviews (1997) 73-78). PNAs specific for a variant EGFR can be introduced into or expressed in a cell as a therapeutic. PNAs have been described, for example, in U.S. Application No. 20040063906.

[00248] Patients to be treated with a compound which targets a variant EGFR include, for example, patients diagnosed with a primary or secondary mutation in their EGFR, patients who initially respond to therapy with a tyrosine kinase

inhibitor, but subsequently fail to respond to the same or similar compound. Alternatively, compounds that target secondary EGFR mutations may be given to cancer patients in combination with compounds that target primary EGFR mutations, for example, gefitinib, as a combination therapy. By combining compounds that target both primary and secondary EGFR mutations, the likelihood of resistance will be reduced.

[00249] Additional EGFR mutations that confer resistance to currently known anti-cancer therapeutics, including but not limited to EGFR tyrosine kinase inhibitors gefitinib, erlotinib and the like, are within the scope of the invention. Resistant EGFR mutants are predicted to have mutants analogous to mutants identified in kinase domains of related tyrosine kinase domain containing proteins that have high homology in this kinase region. Papers describing mutations in analogous proteins include those known in the art for BCR-ABL. See, e.g., Bradford et al. Blood. 2003 Jul 1;102(1):276-83, Epub 2003 Mar 06; Hochhaus et al., Leukemia. 2002 Nov;16(11):2190-6; and Al-Ali et al., Hematol J. 2004;5(1):55-60.

[00250] A mutant EGFR resistant to known EGFR tyrosine kinase inhibitors includes any one or more EGFR polypeptides, or a nucleotide encoding the same, with a non-wild type residue at one or more positions analogous to c-abl (BCR-ABL) residues that confirm an imatinib resistant phenotype. The residues that when mutated in EGFR confer drug resistance include especially those residues from the kinase domain, including but not limited to, e.g., the P-loop and the activation loop, wherein the mutated residues in the EGFR polypeptide are analogous to c-able residues. Contemplated resistant EGFR mutants have non-wild type residues at the amino acids positions that are analogous to at least positions Met 244, Leu 248, Gly 250, Gln 252, Tyr 253, Glu 255, Asp 276, Thr 315, Phe 31 7, Met 351, Glu 355, Phe 359, His 396, Ser 417, and Phe 486 of BCR-ABL, see, for example Table S3C and FIG. 9. These BCL-ABL residues correspond to residues Lys 714, Leu 718, Ser 720, Ala 722, Phe 723, Thr 725, Ala 750, Thr 790, Leu 792, Met 825, Glu 829, Leu 833, His 870, Thr 892, Phe 961, respectively, in EGFR. See, e.g., Table S3C, FIG. 9.

Prognostic Testing

[00251] The methods of the present invention are used as a prognostic indicator of the development of cancer. Alternatively, the methods are used to detect cancer that is present but has not yet been diagnosed or is at a stage that is undetectable. Patients at risk for developing cancer are screened, using the methods of the present invention, for the presence of kinase activity increasing nucleic acid variation in the erbB1 gene. The presence of a variance or variances in the kinase domain of the erbB1 gene indicate the presence or imminent presence of cancer. Thus, the presence of variances in the kinase domain of the erbB1 gene suggest that a patient would benefit from an EGFR targeted treatment. As described herein, an EGFR targeted treatment is preferably treatment with a tyrosine kinase inhibitor.

[00252] In a preferred embodiment of the present invention, a patient is screened for the presence or absence of nucleic acid variances in the kinase domain of the erbB1 gene by obtaining a biological sample. The sample may be any sample from the patient including tissue, e.g., from the tongue, mouth, cheek, trachea, bronchial tube, lungs, etc. or fluid, e.g., from sputum or lung aspirates. Methods of obtaining these biological specimens are well known to those of skill in the art.

[00253] Thus, the invention provides a method for identifying a disease or disorder associated with aberrant mutant EGFR expression or activity in which a test sample is obtained from a subject and mutant EGFR protein or nucleic acid (e.g., mRNA, genomic DNA) is detected, wherein the presence of mutant EGFR protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant mutant EGFR expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (e.g., serum), cell sample, or tissue, especially a tissue biopsy sample.

[00254] Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant mutant EGFR

expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder. Thus, the invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant mutant EGFR expression or activity in which a test sample is obtained and mutant EGFR protein or nucleic acid is detected (e.g., wherein the presence of mutant EGFR protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with mutant EGFR expression or activity).

EXAMPLES

Example 1

Nucleotide Sequence Analysis of Tumor Specimens

[00255] Tumor specimens from initial diagnostic or surgical procedures were collected from patients with NSCLC who were subsequently treated with Gefitinib, under an IRB-approved protocol. Frozen tumor specimens, along with matched normal tissue, were available for four cases, and paraffin-embedded material was used for the remaining specimens. In addition, 25 unselected cases of primary NSCLC (15 bronchioalveolar, 7 adenocarcinoma, and 3 large cell lung cancers), with matched normal tissues, were obtained from the Massachusetts General Hospital tumor bank. For mutational analysis of the entire *EGFR* coding sequence, DNA was extracted from specimens, followed by amplification of all 28 exons, automated sequencing of uncloned PCR fragments, and analysis of electropherograms in both sense and antisense direction for the presence of heterozygous mutations. All sequence variants were confirmed by multiple independent PCR amplifications. Primer sequences and amplification conditions are provided in Supplementary Material. *EGFR* mutations in exons 19 and 21 were also sought in primary tumors of the breast (15 cases), colon (20 cases), kidney (16 cases), and brain (4 cases), along

with a panel of 78 cancer-derived cell lines representing diverse histologies (listed below).

Functional Analysis of Mutant EGFR Constructs

[00256] The L858R and delL747-P753insS mutations were introduced into the full length *EGFR* coding sequence using site-directed mutagenesis and inserted into a cytomegalovirus-driven expression construct (pUSE, Upstate). Cos-7 cells were transfected (Lipofectamine 2000, Invitrogen) using 1 μg of the expression constructs, followed after 18 hrs by replating at 5 x10^4 cells/ well (12-well plates, Costar) in DMEM lacking fetal calf serum. After 16 hrs of serum starvation, cells were stimulated with 10 ng/ml of EGF (SIGMA). To demonstrate Gefitinib inhibition, the drug was added to the culture medium 3 hrs prior to the addition of EGF (30 min stimulation with 100 ng/ml of EGF). Cell lysates were prepared in 100 μL of Laemmli lysis buffer, followed by resolution of proteins on 10% SDS-PAGE, transfer to PVDF membranes, and Western blot analysis using enhanced chemiluminescence reagent (Amersham). Autophosphorylation of EGFR was measured using antibody to phosphotyrosine Y-1068, and comparable protein expression was shown using anti-EGFR antibody (working concentration of 1:1000; Cell Signaling Technology).

MUTATIONAL ANALYSIS

[00257] The polymerase chain reaction was used to amplify the 28 exons comprising the EGFR gene using DNA isolated from primary tumor tissue or tumor-derived cell-lines. Primer pairs used were: Exon 1, CAGATTTGGCTCGACCTGGACATAG (sense) (SEQ ID NO: 513) and CAGCTGATCTCAAGGAAACAGG (antisense) (SEQ ID NO: 514); Exon 2, GTATTATCAGTCAC TAAAGCTCAC (sense) (SEQ ID NO: 515) and CACACTTCAAGTGGAATTCTGC (SEQ ID NO: 516); Exon 3, CTCGTG TGCATTAGGGTTCAACTGG (sense) (SEQ ID NO: 517) and CCTTCTCCGAGGTGGAATTGAGTGAC (antisense) (SEQ ID NO: 518); Exon 4, GCTAATTGCGGGACTCTTGTTCGCAC (sense) (SEQ ID NO: 519) and

TACATGC TTTTCTAGTGGTCAG (antisense) (SEQ ID NO: 520); Exon 5, GGTCTCAAGTGATTCTACAAACCAG (sense) (SEQ ID NO: 521) and CCTTCACCTACTGGTTCACATCTG (antisense) (SEQ ID NO: 522); Exon 6, CATGGT TTGACTTAGTTTGAATGTGG (sense) (SEQ ID NO: 523) and GGATACTAAAGATACTTTGTCAC CAGG(antisense) (SEQ ID NO: 524); Exon 7, GAACACTAGGCTGCAAAGACAGTAAC (sense) (SEQ ID NO: 525) and CCAAGCAAGGCAAACACATCCACC(antisense) (SEQ ID NO: 526); Exon 8, GGAGGATGGAGCC TTTCCATCAC (sense) (SEQ ID NO: 527) and GAAGAGGAAGATGTGTTCCTTTGG (antisense) (SEQ ID NO: 528) ; Exons 9 and 10, GAATGAAGGATGATGTGGCAGTGG (sense) (SEQ ID NO: 529) and CAAAACATCAGCC ATTAACGG (antisense) (SEQ ID NO: 530) ; Exon 11, CCACTTACTGTTCATATAATACAGAG (sense) (SEQ ID NO: 531) and CATGTGAGATAGCATTTGGGAATGC (antisense) (SEQ ID NO: 532) ; Exon 12, CATGACCT ACCATCATTGGAAAGCAG (sense) (SEQ ID NO: 533) and GTAATTTCACAGTTAGGAATC (sense) (SEQ ID NO: 534) ; Exon 13, GTCACCCAAGGTCATGGAGCACAGG (sense) (SEQ ID NO: 535) and CAGAATGC CTGTAAAGCTATAAC (antisense) (SEQ ID NO: 536) ; Exon 14, GTCCTGGAGTCCCAACTCCTTGAC (sense) (SEQ ID NO: 537) and GGAAGTGGCTCTGA TGGCCGTCCTG (antisense) (SEQ ID NO: 538) ; Exon 15, CCAC TCACACACACTAAATATTTTAAG (sense) (SEQ ID NO: 539) and GACCAAAACACCTTAAGTAA CTGACTC (antisense) (SEQ ID NO: 540) ; Exon 16, CCAA TCCAACATCCAGACACATAG (sense) (SEQ ID NO: 541) and CCAGAGCCATAGAAACTTGATCAG (antisense) (SEQ ID NO: 542) ; Exon 17, GTATGGACTATGGC ACTTCAATTGCATGG (sense) (SEQ ID NO: 543) and CCAGAGAACATGGCAACCAGCACAGGAC (antisense) (SEQ ID NO: 544) ; Exon 18, CAAATGAGCTGGCAAGTGCCGTGTC (sense) (SEQ ID NO: 545) and GAGTTT CCCAAACACTCAGTGAAAC (antisense) (SEQ ID NO: 546) or CAAGTGCCGTGTCCTGGCACCCAAGC (sense) (SEQ ID NO: 675) and CCAAACACTCAGTGAAACAAAGAG (antisense) (SEQ ID NO: 676); Exon 19, GCAATATCAGCC TTAGG TGCGGCTC (sense) (SEQ ID NO: 547) and CATAGAAAGTGAACATTTAGGATGTG (antisense) (SEQ ID NO: 548) ; Exon

20, CCATGAGTACGTATTTGAAACTC (sense) (SEQ ID NO: 549) and CATATCC CCATGGC AAACTCTTGC (antisense) (SEQ ID NO: 550) ; Exon 21, CTAACGTTCGCCAG CCATAAGTCC (sense) (SEQ ID NO: 551) and GCTGCGAGCTCACCCAGAATGTCTGG (antisense) (SEQ ID NO: 552) ; Exon 22, GACGGG TCCTGGGGTGATCTGGCTC (sense) (SEQ ID NO: 553) and CTCAGTACAATAGATAGACAGCAATG (antisense) (SEQ ID NO: 684) ; Exon 23, CAGGACTACAGAAATGTAGGTTTC (sense) (SEQ ID NO: 555) and GTGCCTG CCTTAAGTAATGTGATGAC (antisense) (SEQ ID NO: 556) ; Exon 24, GACTGG AAGTGTCGCA TCACCAATG (sense) (SEQ ID NO: 557) and GGTTTAATAATGCGATCTGGGACAC (antisense) (SEQ ID NO: 558) ; Exon 25, GCAGCTATAATTTAGAGAACCAAGG (sense) (SEQ ID NO: 559) and GGTT AAAATTGACTTC ATTTCCATG (antisense) (SEQ ID NO: 560) ; Exon 26, CCTAGTTGCTCTAAA ACTAACG (sense) (SEQ ID NO: 561) and CTGTGAGGCGTGACAGCCGTGCAG (antisense) (SEQ ID NO: 562) ; Exon 27, CAACCTACTAATCAG AACCAGCATC (sense) (SEQ ID NO: 563) and CCTTCACTGTGTCTGC AAATCTGC (antisense) (SEQ ID NO: 564) ; Exon 28, CCTGTCATAAGTCTCCTTGTTGAG (sense) (SEQ ID NO: 565) and CAGTCTGTGGGTCTAAG AGCTAATG (antisense) (SEQ ID NO: 566) . Annealing temperatures were 58°C (exons 1,3, 4, 7-10, 12-25, 27, and 28), 56°C (exons 2, 5, 6, and 26), or 52°C (exon 11).

[00258] Nested PCR amplification of DNA extracted from archival tumor tissue was performed as follows. An initial PCR for exons 2, 5, 6, 7, 11, 12, 14, 16, 18, 19, 20, 21, 23, 24, 25, 26, and 27 was generated using primers and conditions described above. Subsequently, 2 μl of this reaction was amplified in a secondary PCR using the following internal primer pairs: Exon 2, CAGGAATGGGTGAGTCTCTGTGTG (sense) (SEQ ID NO: 567) and GTGGAATTCTGCCCAGGCCTTTC (antisense) (SEQ ID NO: 568) ; Exon 5, GATTCTACAAACCA GCCAGCCAAAC (sense) (SEQ ID NO: 569) and CCTACTGGTTCACATCTGACCCTG (antisense) (SEQ ID NO: 570) ; Exon 6, GTTTGAATGTGGTTTCGTTGGAAG (sense) (SEQ ID NO: 571) and CTTTGTCACCAGG CAGAGG GCAATATC (antisense) (SEQ ID NO: 572) ; Exon

7, GACAGTAACTTGGGCTTTCTGAC (sense) (SEQ ID NO: 573) and CATCCACCCAAAGACTCTCCAAG (antisense) (SEQ ID NO: 574) ; Exon11, CTGTTCATA TAATAC AGAGTCCCTG (sense) (SEQ ID NO: 575) and GAGAGATGCAGGAGCTCTGTGC (antisense) (SEQ ID NO: 576) ; Exon12, GCAGTTTGTAGTCAATCAAAGGTGG (sense) (SEQ ID NO: 577) and GTAATTTAAATGGGAAT AGCCC (antisense) (SEQ ID NO: 578) ; Exon14, CAACTCCTTGACCATTACCTCAAG (sense) (SEQ ID NO: 579) and GATGGCCGTCCTGCCCACACAGG (antisense) (SEQ ID NO: 580) ; Exon16, GAGTAGTTTAGCA TATATTGC (sense) (SEQ ID NO: 581) and GACAGTCAGAAATGCAGGAAAGC (antisense) (SEQ ID NO: 582) ; Exon18, CAAGTGCCGTGTCCTGGCACCCAAGC (sense) (SEQ ID NO: 583) and CCAAACACTCA GTGAAACAAAGAG (antisense) (SEQ ID NO: 584) or GCACCCAAGCCCATGCCGTGGCTGC (sense) (SEQ ID NO: 677)and GAAACAAAGAGTAAAGTAGATGATGG (antisense) (SEQ ID NO: 678); Exon 19, CCTTAGGTGCGGCTCCACAGC (sense) (SEQ ID NO: 585) and CATTTAGGATGTGGAGATGAGC (antisense) (SEQ ID NO: 586) ; Exon 20, GAAACTCAAG ATCGCATTCATGC (sense) (SEQ ID NO: 587) and GCAAACTCTTGCTATCCCAGGAG (antisense) (SEQ ID NO: 588) ; Exon 21, CAGCCATAAGTCCTCGACGTGG (sense) (SEQ ID NO: 589) and CATCCTCCCCT GCATGTGTTAAAC (antisense) (SEQ ID NO: 590); Exon 23, GTAGGTTTCTAAACATCAAGAAAC (sense) (SEQ ID NO: 591) and GTGATGACATTTCTCCAGGGATGC (antisense) (SEQ ID NO: 592) ; Exon 24, CATCACCA ATGCCTTCTTTAAGC (sense) (SEQ ID NO: 593) and GCTGGAGGGTTTAATAATGCGATC (antisense) (SEQ ID NO: 594) ; Exon 25, GCAAACACACAGGCACCTGCTGGC (sense) (SEQ ID NO: 595) and CATTTC CATGTGAGTTTCACTAGATGG (antisense) (SEQ ID NO: 596); Exon 26, CACCTTCACAATATACCCTCCATG (sense) (SEQ ID NO: 679) and GACAGCCGTGCAGGGAAAAACC (antisense) (SEQ ID NO: 680); Exon 27, GAACCAGCATCTCAAGGAGATCTC (sense) (SEQ ID NO: 681) and GAGCACCTGGCTTGGACACTGGAG (antisense) (SEQ ID NO: 682).

[00259] Nested PCR amplifications for the remaining exons consisted of primary PCR using the following primers. Exon 1, GACCGGACGACAGGCCACCTCGTC (sense) (SEQ ID NO: 597) and GAAGAACGAAACGTCCCGTTCCTCC (antisense) (SEQ ID NO: 598) ; Exon 3, GTTGAGCACT CGTGTGCATTAGG (sense) (SEQ ID NO: 599) and CTCAGTGCACGTGTACTGGGTA (antisense) (SEQ ID NO: 600) ; Exon 4, GTTCACTGGGCTAATTGCGGGACTCTTGTTCGCAC (sense) (SEQ ID NO: 601) and GGTA AATACATGCTTTTCTAGTGGTCAG (antisense) (SEQ ID NO: 602) ; Exon 8, GGAGGATGGA GCCTTTCCATCAC (sense) (SEQ ID NO: 603) and GAAGAGGAAGATGTGTTCCTTTGG (antisense) (SEQ ID NO: 604) ; Exon 9, GAATGAAGGATGATGTGGCAGTGG (sense) (SEQ ID NO: 605) and GTATGTGTGAAGGAG TCACTGAAAC (antisense) (SEQ ID NO: 606) ; Exon 10, GGTGAGTCACAGGTTCAGTTGC (sense) (SEQ ID NO: 607) and CAAAACATCAGCCATTAACGG (antisense) (SEQ ID NO: 608) ; Exon 13, GTAGCCAGCATGTC TGTGTCAC (sense) (SEQ ID NO: 609) and CAGAATGCCTGTAAAGCTATAAC (antisense) (SEQ ID NO: 610) ; Exon 15, CATTTGGCTTTCCCCACTCACAC (sense) (SEQ ID NO: 611) and GACCAAAACACCTTAA GTAACTGACTC (antisense) (SEQ ID NO: 612) ; Exon 17, GAAGCTACATAGTGTCTCACTTTCC (sense) (SEQ ID NO: 613) and CACAACTGCTAATGGCCCGTTCTCG (antisense) (SEQ ID NO: 614); Exon 22, GAGCAGCCCTGAACTCCGTCAGACTG (sense) (SEQ ID NO: 683) and CTCAGTACAATAGATAGACAGCAATG (antisense) (SEQ ID NO: 684); Exon 28a GCTCC TGCTCCCTGTCATAAGTC (sense) (SEQ ID NO: 615) and GAAGTCCTGCTGGTAGTCAGGGTTG (antisense) (SEQ ID NO: 616) ; Exon 28b, CTGCAGTGGGCAACCCCGAGTATC (sense) (SEQ ID NO: 617) and CAGTC TGTGGGTCTAAGAGCTAATG (antisense) (SEQ ID NO: 618) . Secondary PCR amplification was carried out using primer pairs: Exon 1, GACAGGCCACCTCGTCGGCGTC (sense) (SEQ ID NO: 619) and CAGCTGATCTCAAGGAAACAGG (antisense) (SEQ ID NO: 620) ; Exon 3, CTCGTG TGCATTA GGGTTCAACTGG (sense) (SEQ ID NO: 621) and CCTTCTCCGAGGTGGAATTGAGTGAC (antisense) (SEQ ID NO: 622) ; Exon 4,

GCTAATTGCGGGACTCTTGTTCGCAC (sense) (SEQ ID NO: 623) and TACATGCTTT TCTAGTGGTCAG (antisense) (SEQ ID NO: 624) ; Exon 8, CCTTTCCATCACCCCTCAAGAGG (sense) (SEQ ID NO: 625) and GATGTGTTCCTTTGGAGGTGGCATG (antisense) (SEQ ID NO: 626) ; Exon 9, GATGTGG CAGTGGCGGTTCCGGTG (sense) (SEQ ID NO: 627) and GGAGTCACTGAAACAAACAACAGG (antisense) (SEQ ID NO: 628) ; Exon 10, GGTTCAGTTGCTTGTATAAAG (sense) (SEQ ID NO: 629) and CCATTAACGGT AAAATTTCAGAAG (antisense) (SEQ ID NO: 630) ; Exon 13, CCAAGGTCATGGAGCACAGG (sense) (SEQ ID NO: 631) and CTGTAAAGCTATAACAACAACCTGG (antisense) (SEQ ID NO: 632) ; Exon 15, CCACTCACA CACACTAAATATTTTAAG (sense) (SEQ ID NO: 633) and GTAACTGACTCAAATACAAACCAC (antisense) (SEQ ID NO: 634) ; Exon 17, GAAGCTACATAGTGTCTCACTTTCC (sense) (SEQ ID NO: 635) and CACAA CTGCTAATGGCCCGTTCTCG (antisense) (SEQ ID NO: 636) ; Exon 22, GACGGGTCCTGGGGTGATCTGGCTC (sense) (SEQ ID NO: 685) and CTCAGTACAATAGATAGACAGCAATG (antisense) (SEQ ID NO: 686); Exon 28a, CCTGTCATAAG TCTCCTTGTTGAG (sense) (SEQ ID NO: 637) and GGTAGTCAGGGTTGTCCAGG (antisense) (SEQ ID NO: 638) ; Exon 28b, CGAGTATCTCAACACTGTCCAGC (sense) (SEQ ID NO: 639) and CTAAGAGCTAATGCGGGC ATGGCTG (antisense) (SEQ ID NO: 640) . Annealing temperature for exon 1 amplifications was 54°. Annealing temperatures for both primary and secondary amplifications were 58°C (exons 3, 4, 7-10, 12-17, 19-25, 27, and 28), 56°C (exons 2, 5, 6, and 26), or 52°C (exons 11 and 18).

[00260] PCR amplicons were purified using exonuclease I (United States Biochemical, Cleveland, OH), and shrimp alkaline phosphatase (United States Biochemical, Cleveland, OH) prior to sequencing. Purified DNA was diluted and cycle-sequenced using the ABI BigDye Terminator kit v1.1 (ABI, Foster City, CA) according to manufacturer's instructions. Sequencing reactions were electrophoresed on an ABI3100 genetic analyzer. Electropherograms were analyzed in both sense and antisense direction for the presence of mutations, using Sequence Navigator software in combination with Factura to mark heterozygous positions. All sequence variants

were confirmed in multiple independent PCR amplifications and sequencing reactions.

Cancer-Derived Cell Lines:

[00261] A panel of 14 lung cancer-derived cell lines was analyzed for *EGFR* mutations. These were derived from tumors of NSCLC (N=5), small cell lung cancer (N=6), adenosquamous (N=1), bronchial carcinoid (N=1), and unknown histology (N=1). Specific cell lines were: NCI-H460, NCI-522, HOP-92, NCIH841, NCIH734, NCIH2228, NCIH596, NCIH727, NCIH446, NCIH1781, NCIH209, NCIH510, NCIH82, NCIH865. In addition, 64 cancer-derived cell lines were screened for mutations in exons 19 and 21. These represented the following histologies: breast cancer (BT549, BT483, UACC893, HS467T, HS578T, MCF7, MCF7-ADR, MDA-MB-15, MDA-MB-175, MDA-MB-231, MDA-MB-415, MDA-MB-436, MDA-MB-453, MDA-MB-468, T47D), ovarian cancer (ES-2, IGROV-1, MDAH2774, OV1063, OVCAR3, OVCAR4, OVCAR5, SKOV3, SW626), CNS cancers (SF-295, SNB-19, U-251, CCF-STTG1, SW-1088, SW-1783, T98G, M059K, A172, SK-N-DZ, SK-N-MC), leukemia (CCRF-CEM, K562, MOLT-4, RPMI8226, SR), prostate cancer (DU-145, PC-3), colon cancer (COLO-205, HCT-116, HCT-15, HT-29, SW-620), renal cancer(786-O, ACHN, CAKI-1, SN-12C, UO-31), melanoma (LOX-IMVI, M14, SKMEL2, UACC-62), osteosarcoma (SAOS-2), and head and neck cancers (O11, O13, 019, O28, O22, O29, O12). The head and neck cancer cell-lines were provided by Dr. James Rocco, Massachusetts General Hospital/Massachusetts Eye and Ear Infirmary. All other cell-lines are available through the American Type Culture Collection (Manassas, VA).

[00262] Genomic DNA was isolated from snap-frozen tumor specimens. Tumor specimens were first crushed to a fine powder using a pre-chilled and sterilized mortar and pestle. Tumor tissue was immediately transferred into a DNA extraction solution consisting of 100mM sodium chloride, 10mM Tris pH7.5, 25mM EDTA (disodium ethylenediamine tetraacetate) pH8.0, and 0.5% (w/v) sodum dodecyle sulfate, and 100µg/ml fresh proteinase K and incubated overnight at 37°C or for 3 hours at 50°C. DNA was then extracted using standard phenol-chloroform

methods, ethanol precipitated, washed with 70fi ethanol, air-dried and resuspended in TE buffer. The DNA concentration was determined spectrophotometrically. Exons 19 and 21 of human EGFR were amplified by the polymerase chain reaction using the following primer pairs: Exon19 sense primer, 5'-GCAATATCAGCCTTAGGTGCGGCTC-3' (SEQ ID NO: 505); Exon 19 antisense primer, 5'-CATAGAA AGTGAACATTTAGGATGTG-3' (SEQ ID NO: 506); Exon 21 sense primer, 5'-CTAACGTTCG CCAGCCATAAGTCC-3' (SEQ ID NO: 507); Exon21 antisense primer, 5'- GCTGCGAGCTCACCCAG AATGTCTGG-3' (SEQ ID NO: 508). For each sample, 20ng of genomic DNA was amplified in a PCR reaction consisting of 1X Expand Long Template buffer 1 (Roche, Mannhein Germany), 50μM sequencing grade dATP (Amersham Biosciences, Cleveland OH), 50μM sequencing grade dCTP (Amersham Biosciences, Cleveland OH), 50μM sequencing grade dGTP (Amersham Biosciences, Cleveland OH), 50μM sequencing grade dTTP (Amersham Biosciences, Cleveland OH), 0.2μM sense primer, 0.2μM antisense primer, 1.25 units Expand Long Template enzyme mix (Taq DNA polymerase/Tgo DNA polymerase) (Roche, Mannhein Germany) that has been preincubated for 5 minutes on ice with 1/6 volume of TaqStart Antibody (1.1μg/μl) (Clontech, Palo Alto, CA) and water to final volume of 25μl. Each series of amplifications also includes a negative control for which the DNA template is omitted. PCR cycling conditions for both exons were 95°C for 2 min followed by 40 cycles of 95°C for 30 s, 58°C for 30 s and 72°C for 45 sec; and a final extension of 72°C for 10 min followed by holding at 4°C on an MJ-Research PTC-200 or PTC-225 thermal-cycler (MJ-Research, Waltham MA).

[00263] PCR products were resolved by electrophoresis through a 0.8% agarose gel to ensure amplification from patient material and no amplification in the negative control. PCR products were purified prior to sequencing by mixing 10μl each PCR amplicon with 0.5μl exonuclease I (10U/μl) (United States Biochemical, Cleveland, OH), and 1μl shrimp alkaline phosphatase (1U/μl) (United States Biochemical, Cleveland, OH) and incubating at 37°C for 20 minutes followed by inactivation at 80°C for 15 minutes on a termal-cycler (MJ-Research, Waltham, MA). Purified DNA was diluted in water, according to the intensity of the amplicon, and

cycle-sequencing was performed using the ABI BigDye Terminator kit v1.1 (Applied Biosystems, Foster City, CA) according to manufacturer's instructions. Cycle-sequencing was performed on an MJ-Research thermal-cycler using the following cycling conditions: Primers used for sequencing were: Exon19 sense primer, 5'-GCAATATCAGCCTTAGGTGCGGCTC-3' (SEQ ID NO: 505); Exon 19 antisense primer, 5'- CATAG AAAGTGAACATTTAGGATGTG-3' (SEQ ID NO: 506); Exon21 sense primer, 5'-CTAACGTTCGCCAG CCATAAGTCC-3' (SEQ ID NO: 507) or 5'-CGTGGAGAGGCTCAGAGCCTGGCATG-3' (SEQ ID NO: 687); Exon 21 antisense primer, 5'-GCTGCGAGCTCACCCAGAATGTCTGG-3' (SEQ ID NO: 508). Sequencing reactions were electrophoresed on an ABI3100 genetic analyzer (Applied Biosystems, Foster City, CA). Factura and Sequence Navigator (Applied Biosystems, Foster City, CA) software programs were used to mark potential heterozygous positions and display them for evaluation. Nucleotide positions at which the height of the secondary peak was greater than, or equal to, 30% the height of the primary peak were marked as heterozygous and were confirmed by analysis of both sense and antisense reads. Samples with sequence indicative of the presence of a mutation were re-amplified and sequenced for confirmation.

Position of Primers Used in Sequence Analysis Relative to Exons 19 and 21

Intronic primers are shown in lower case and underlined.

Intronic Sequence is Shown in Lowercase.

Exonic Sequence is Shown in Uppercase.

[00264] EGFR Exon 19 (5'-3') (SEQ ID NO: 641)

gcaatatcagccttaggtgcggctccacagccccagtgtccctcaccttcggggtgcatcgc
tggtaacatccacccagatcactgggcagcatgtggcaccatctcacaattgccagttaacg
tcttccttctctctctgtcatagGGACTCTGGATCCCAGAAGGTGAGAAAGTTAAAATTCCC
GTCGCTATCAAGGAATTAAGAGAAGCAACATCTCCGAAAGCCAACAAGGAAATCCTCGATgt
gagtttctgctttgctgtgtgggggtccatggctctgaacctcaggcccaccttttctcatg
tctggcagctgctctgctctagaccctgctcatctccacatcctaaatgttcactttctatg

[00265] EGFR Exon 21 (5'-3') (SEQ ID NO: 642) or (SEQ ID NO: 687)

ctaacgttcgccagc cataagtcctcgacgtggagaggctcagagcctggcatgaacatgac
cctgaattcggatgcagagcttcttcccatgatgatctgtccctcacagcagggtcttctct
gtttcagGGCATGAACTACTTGGAGGACCGTCGCTTGGTGCACCGCGACCTGGCAGCCAGGA
ACGTACTGGTGAAAACACCGCAGCATGTCAAGATCACAGATTTTGGGCTGGCCAAACTGCTG
GGTGCGGAAGAGAAAGAATACCATGCAGAAGGAGGCAAAgtaaggaggtggctttaggtcag
ccagcattttcctgacaccagggaccaggctgccttcccactagctgtattgtttaacacat
gcaggggaggatgctctccagacattctgggtgagctcgcagc

RESULTS

Clinical Characteristics of Gefitinib Responders

[00266] Patients with advanced, chemotherapy-refractory NSCLC have been treated with single agent Gefitinib since 2000 at Massachusetts General Hospital. A total of 275 patients were treated, both prior to its approval on May 2003 by the FDA, as part of a compassionate use expanded access program, and following that date using commercial supply. During this period, 25 patients were identified by clinicians as having significant clinical responses. A significant clinical response was defined either as a partial response using RECIST criteria for patients with measurable disease, or for patients whose tumor burden could not be quantified using these criteria, an evaluable response was assessed by two physicians. Table 1 shows clinical characteristics of 9 cases for whom tumor specimens obtained at the time of initial diagnosis were available. For the other Gefitinib-responders, tissue was not available, most commonly because diagnostic specimens were limited to cytology from needle aspirates. As a group, the 9 patients experienced substantial benefit from Gefitinib. The median survival from the start of drug treatment is in excess of 18 months, and the median duration of therapy is greater than 16 months. Consistent with previous reports, Gefitinib-responders have a high prevalence of female sex, absence of smoking history, and tumors with bronchioalveolar histology (11, 12). Case 6 is representative of the Gefitinib-responsive cohort. This patient is a 32 year-old man, without smoking history, who presented with multiple brain lesions and disease in the right lung diagnosed as bronchioalveolar carcinoma. He was treated with whole brain radiotherapy, followed by a series of chemotherapy regimens to which his tumor did

not respond (carboplatin and gemcitabine; docetaxel; vinorelbine). With a declining functional status and progressive lung tumor burden, he started therapy with 250 mg per day of Gefitinib. His shortness of breath promptly improved and a lung CT scan 6 weeks after initiation of treatment revealed the dramatic improvement shown in Figure 1.

EGFR Mutations in Gefitinib Responders

[00267] We hypothesized that cases of NSCLC with striking responses to Gefitinib might harbor somatic mutations in *EGFR*, indicating an essential role played by this growth factor signaling pathway in these tumors. To search for such mutations, we first tested for rearrangements within the extracellular domain of EGFR that are characteristic of gliomas (15): none were detected. We therefore sequenced the entire coding region of the gene using PCR-amplification of individual exons. Heterozygous mutations were observed in 8/9 cases, all of which were clustered within the kinase domain of EGFR (Table 2 and Figure 2). Four tumors had in-frame deletions removing amino acids 746-750 (delE756-A750; case 1), 747 to 750 (delL747-T751insS; case 2), and 747 to 752 (delL747-P753insS; cases 3 and 4). The latter two deletions were associated with the insertion of a serine residue, resulting from the generation of a novel codon at the deletion breakpoint. Remarkably, these four deletions were overlapping, with the deletion of four amino acids (leucine, arginine, glutamic acid and alanine, at codons 747 to 750) within exon 19 shared by all cases (see Figure 4a). Another three tumors had amino acid substitutions within exon 21: leucine to arginine at codon 858 (L858R; cases 5 and 6), and leucine to glutamine at codon 861 (L861Q; case 7). The L861Q mutation is of particular interest, since the same amino acid change in the mouse *egfr* gene is responsible for the Dark Skin (dsk5) trait, associated with altered EGFR signaling (18). A fourth missense mutation in the kinase domain resulted in a glycine to cysteine substitution at codon 719 within exon 18 (G719C; case 8). Matched normal tissue was available for cases 1, 4, 5 and 6, and showed only wild-type sequence, indicating that the mutations had arisen somatically, during tumor formation. No mutations were

observed in seven cases of NSCLC that failed to respond to Gefitinib ($P=0.0007$; 2-sided Fisher's exact test).

Prevalence of Specific EGFR Mutations in NSCLC and Other Cancer Types

[00268] Unlike gliomas, in which rearrangements affecting the EGFR extracellular domain have been extensively studied (15), the frequency of *EGFR* mutations in NSCLC has not been defined. We therefore sequenced the entire coding region of the gene in 25 primary cases of NSCLC unrelated to the Gefitinib study, including 15 with bronchioalveolar histology, which has been associated with Gefitinib-responsiveness in previous clinical trials (11, 12). Heterozygous mutations were detected in two bronchioalveolar cancers. Both cases had in-frame deletions in the kinase domain identical to those found in Gefitinib responders, namely delL747-P753insS and delE746-A750 (Table 2). Given the apparent clustering of *EGFR* mutations, we sequenced exons 19 and 21 in a total of 55 primary tumors and 78 cancer-derived cell lines, representing diverse tumor types (see Supplementary Material). No mutations were detected, suggesting that these arise only in a subset of cancers, in which EGFR signaling may play a critical role in tumorigenesis.

Increase in EGF-Induced Activation and Gefitinib Inhibition of Mutant EGFR Proteins

[00269] To study the functional properties encoded by these mutations, the L747-S752insS deletion and the L858R missense mutants were expressed in cultured cells. Transient transfection of wild-type and mutant constructs into Cos-7 cells demonstrated equivalent expression levels, indicating that the mutations do not affect protein stability. EGFR activation was quantified by measuring phosphorylation of the tyrosine ^{1068} residue, commonly used as a marker of receptor autophosphorylation (19). In the absence of serum and associated growth factors, neither wild-type nor mutant EGFR demonstrated autophosphorylation (Figure 3a). However, addition of EGF led to a 2-3 fold increase in receptor activation for both the missense and deletion *EGFR* mutants, compared with the wild-type receptor. Moreover, whereas

normal EGFR activation was downregulated after 15 min, consistent with receptor internalization, the two mutant receptors demonstrated continued activation for up to 3 hrs (Figure 3a). Similar results were obtained with antibodies measuring total EGFR phosphorylation following addition of EGF (not shown).

[00270] Since 7/8 EGFR kinase mutations reside near the ATP cleft, which is targeted by Gefitinib, we determined whether the mutant proteins have altered sensitivity to the inhibitor. EGF-induced receptor autophosphorylation was measured in cells pretreated with variable concentrations of Gefitinib. Remarkably, both mutant receptors displayed increased sensitivity to inhibition by Gefitinib. Wild-type EGFR had an IC_{50} of 0.1μM and showed complete inhibition of autophosphorylation at 2μM Gefitinib, whereas the two mutant proteins had an IC_{50} of 0.015μM and abrogation of autophosphorylation at 0.2μM (Figure 3b). This difference in drug sensitivity may be clinically relevant, since pharmacokinetic studies indicate that daily oral administration of 400-600 mg of Gefitinib results in a mean steady-state trough plasma concentration of 1.1-1.4 μM, while the currently recommended daily dose of 250 mg leads to a mean trough concentration of 0.4 μM (20).

Example 2

[00271] Tumor cells harboring mutations within the kinase domain of the EGFR, and are therefore sensitive to growth inhibition by gefitinib treatment, can undergo "second-site" mutations, also within the kinase domain, that confer resistance to gefitinib but are still "activating" in the sense that they exhibit increased EGFR signaling relative to wild-type EGFR. Such gefitinib-resistant mutants are generated from two sporadic human NSCLC cell lines namely NCI-1650 and NCI-1975. Each cell line contains a heterozygous mutation with the kinase domain of EGFR, and is, therefore, expected to be sensitive to gefitinib. The EGFR mutation in NCI-1650 consists of an in-frame deletion of 15 nucleotides at position 2235-2249 (delLE746-A750) within exon 19, while NCI-1975 has a missense mutation within exon 21 that substitutes a G for T at nucleotide 2573 (L858R). The L858R mutation in NCI-H1975

has been shown herein to be activating and to confer increased sensitivity to gefitinib in vitro.

[00272] Gefitinib-resistant cell lines, derived from both NCI-1650 and NCI-1975 are isolated, following random chemical mutagenesis using EMS (ethyl methanesulfonate) followed by culture in gefitinib-supplemented medium to select for the outgrowth of resistant clones. Subcultivation of individual clones is followed by nucleotide sequence determination of the EGFR gene following specific PCR-mediated amplification of genomic DNA corresponding to the EGFR kinase domain.

[00273] A variation of this strategy involves the serial passage of these two cell lines in the presence of gradually increasing concentrations of gefitinib over a course of several weeks or months in order to select for the spontaneous acquisition of mutations within the EGFR gene that confer resistance to gefitinib. Selected cells (that continue to proliferate at relatively high gefitinib concentration) are isolated as colonies, and mutations are identified as described above.

Example 3

[00274] To determine whether mutation of receptor tyrosine kinases plays a causal role in NSCLC, we searched for somatic genetic alterations in a set of 119 primary NSCLC tumors, consisting of 58 samples from Nagoya City University Hospital in Japan and 61 from the Brigham and Women's Hospital in Boston, Massachusetts. The tumors included 70 lung adenocarcinomas and 49 other NSCLC tumors from 74 male and 45 female patients, none of whom had documented treatment with EGFR kinase inhibitors.

[00275] As an initial screen, we amplified and sequenced the exons encoding the activation loops of 47 of the 58 human receptor tyrosine kinase genes (*) (Table S1) from genomic DNA from a subset of 58 NSCLC samples including 41 lung adenocarcinomas. Three of the tumors, all lung adenocarcinomas, showed heterozygous missense mutations in *EGFR* not present in the DNA from normal lung tissue from the same patients (Table S2; S0361, S0388, S0389). No mutations were detected in amplicons from other receptor tyrosine kinase genes. All three tumors had the same *EGFR* mutation, predicted to change leucine ("L") at position 858 to

arginine ("R") (FIG. 6A; CTG→CGG; "L858R"), wherein all numbering refers to human EGFR.

[00276] We next examined exons 2 through 25 of *EGFR* in the complete collection of 119 NSCLC tumors. Exon sequencing of genomic DNA revealed missense and deletion mutations of *EGFR* in a total of 16 tumors, all within exons 18 through 21 of the kinase domain. All sequence alterations in this group were heterozygous in the tumor DNA; in each case, paired normal lung tissue from the same patient showed wild-type sequence, confirming that the mutations are somatic in origin. The distribution of nucleotide and protein sequence alterations, and the patient characteristics associated with these abnormalities, are summarized in Table S2.

[00277] Substitution mutations G719S and L858R were detected in two and three tumors, respectively. The "G719S" mutation changes the glycine (G) at position 719 to serine (S) (FIG. 6B). These mutations are located in the GXGXXG motif (SEQ ID NO:490) of the nucleotide triphosphate binding domain or P-loop and adjacent to the highly conserved DFG motif in the activation loop (*52*), respectively. *See, e.g.*, FIG. 7. The mutated residues are nearly invariant in all protein kinases and the analogous residues (G463 and L596) in the B-Raf protein serine-threonine kinase are somatically mutated in colorectal, ovarian and lung carcinomas (*41, 53*) (FIG. 6A, 6B).

[00278] We also detected multiple deletion mutations clustered in the region spanning codons 746 to 759 within the kinase domain of EGFR. Ten tumors carried one of two overlapping 15-nucleotide deletions eliminating *EGFR* codons 746 to 750, starting at either nucleotide 2235 or 2236 (Del-1; FIGS. 6C and 8C; Table S2). *EGFR* DNA from another tumor displayed a heterozygous 24-nucleotide gap leading to the deletion of codons 752 to 759 (Del-2; FIG. 6C). Representative chromatograms are shown in FIGS. 8A-8F.

[00279] The positions of the substitution mutations and the Del-1 deletion in the three-dimensional structure of the active form of the EGFR kinase domain (*54*) are shown in FIG. 7. Note that the sequence alterations cluster around the active site of the kinase, and that the substitution mutations lie in the activation loop and glycine-

rich P-loop, structural elements known to be important for autoregulation in many protein kinases (*52*).

[00280] Two additional *EGFR* mutations in two different tumor types have been identified. Namely, we have identified the *EGFR* mutation G857V in Acute Myelogenous Leukemia (AML) and the *EGFR* mutation L883S in a metastatic sarcoma. The "G857V" mutation has the glycine (G) at position 857 substituted with a valine (V), while the "L883S" mutation has the leucine (L) at position 883 substituted with a serine (S). These findings suggest that mutations in EGFR occur in several tumor types and, most importantly, that EGFR inhibitors would be efficacious in the treatment of patients harboring such mutations. This expands the use of kinase inhibitors such as, *e.g.*, the tyrosine kinase inhibitors gefitinib (marketed as Iressa™), erlotinib (marketed as Tarceva™), and the like in treating tumor types other than NSCLC.

[00281] The *EGFR* mutations show a striking correlation with the differential patient characteristics described in Japanese and U.S. patient populations. As noted above, clinical trials reveal significant variability in the response to the tyrosine kinase inhibitor gefitinib (Iressa™), with higher responses seen in Japanese patients than in a predominantly European-derived population (27.5% vs. 10.4%, in a multi-institutional phase II trial) (*48*); and with partial responses seen more frequently in the U.S. in women, non-smokers, and patients with adenocarcinomas (*49-51*). We show that *EGFR* mutations were more frequent in adenocarcinomas (15/70 or 21%) than in other NSCLCs (1/49 or 2%); more frequent in women (9/45 or 20%) than in men (7/74 or 9%), and more frequent in the patients from Japan (15/58 or 26%, and 14/41 adenocarcinomas or 32%) than in those from the US (1/61 or 2%, and 1/29 adenocarcinomas or 3%). The highest fraction of *EGFR* mutations was observed in Japanese women with adenocarcinoma (8/14 or 57%). Notably, the patient characteristics that correlate with the presence of *EGFR* mutations appear to be those that correlate with clinical response to gefitinib treatment.

[00282] To investigate whether *EGFR* mutations might be a determinant of gefitinib sensitivity, pre-treatment NSCLC samples were obtained from 5 patients who responded and 4 patients who progressed during treatment with gefitinib, where

these patients were identified out of more than 125 patients treated at the Dana-Farber Cancer Institute either on an expanded access program or after regulatory approval of gefitinib (*49*). Four of the patients had partial radiographic responses ($\geq$ 50% tumor regression in a CT scan after 2 months of treatment) while the fifth patient experienced dramatic symptomatic improvement in less than two months. All of the patients were from the United States and were Caucasian.

[00283] While sequencing of the kinase domain (exons 18 through 24) revealed no mutations in tumors from the four patients whose tumors progressed on gefitinib, all five tumors from gefitinib-responsive patients harbored *EGFR* kinase domain mutations. The Chi-squared test revealed the difference in *EGFR* mutation frequency between gefitinib responders (5/5) and non-responders (0/4) to be statistically significant with $p = 0.0027$, while the difference between the gefitinib-responders and unselected U.S. NSCLC patients (5/5 vs. 1/61) was also significant with $p < 10^{-12}$ (*). The *EGFR* L858R mutation, previously observed in the unselected tumors, was identified in one gefitinib-sensitive lung adenocarcinoma (FIG. 6A; Table S3A, IR3T). Three gefitinib-sensitive tumors contained heterozygous in-frame deletions (Fig. 6C and Tables S3A and S3B, Del-3 in two cases and Del-4 in one) and one contained a homozygous in-frame deletion (Fig. 6C and Tables S3A and S3B, Del-5). Each of these deletions was within the codon 746 to 753 region of *EGFR* where deletions were also found in unselected tumors. Each of these three deletions is also associated with an amino acid substitution (Tables S3A-S3C). In all four samples where matched normal tissue was available, these mutations were confirmed as somatic.

[00284] Example 3A: Primer design

[00285] The cDNA sequences of receptor tyrosine kinases were obtained from GenBank (accession numbers listed in Table S1), and were to the human genome assembly (http://genome.ucsc.edu) using the BLAT alignment to identify exon/intron boundaries. External gene specific primer pairs were designed to amplify exon sequences and at least 250 bp of flanking intronic sequence or adjacent exonic sequence on each side using the Primer3 program (http://frodo.wi.mit.edu/primer3/primer3_code.html). The resulting predicted

amplicons were then used to design internal primers flanking the exon (generally greater than 50 bp from the exon/intron boundary) and containing appended M13 forward or reverse primer tails. These nested primer sets were tested for appropriate amplicon size and high-quality sequence from control DNA. Amplicons encompassing exons encoding the receptor tyrosine kinase activation loop of 47 tyrosine kinases were amplified and sequenced in a set of 58 primary lung cancer samples from Nagoya City University Medical School. In addition, amplicons covering the full length EGFR were also amplified.

[00286] Example 3B: PCR and sequencing methods for genomic DNA

[00287] Tyrosine kinase exons and flanking intronic sequences were amplified using specific primers in a 384-well format nested PCR setup. Each PCR reaction contained 5 ng of DNA, 1X HotStar Buffer, 0.8 mM dNTPs, 1 mM MgCl2, 0.2U HotStar Enzyme (Qiagen, Valencia, CA), and 0.2 μM forward and reverse primers in a 10 μL reaction volume. PCR cycling parameters were: one cycle of 95°C for 15 min, 35 cycles of 95°C for 20s, 60°C for 30s and 72°C for 1 min, followed by one cycle of 72°C for 3 min.

[00288] The resulting PCR products were purified by solid phase reversible immobilization chemistry followed by bi-directional dye-terminator fluorescent sequencing with universal M13 primers. Sequencing fragments were detected via capillary electrophoresis using ABI Prism 3700 DNA Analyzer (Applied Biosystems, Foster City, CA). PCR and sequencing were performed by Agencourt Bioscience Corporation (Beverly, MA).

[00289] Example 3B: Sequence analysis and validation

[00290] Forward (F) and reverse (R) chromatograms were analyzed in batch by Mutation Surveyor 2.03 (SoftGenetics, State College, PA), followed by manual review. High quality sequence variations found in one or both directions were scored as candidate mutations. Exons harboring candidate mutations were reamplified from the original DNA sample and re-sequenced as above.

[00291] Example 3C: Patients

[00292] Lung tumor specimens were obtained from patients with non-small cell lung cancer treated at Nagoya City University Hospital and the Brigham and

Womens's Hospital (unselected Japanese tumors and gefitinib-treated U.S. tumors, respectively) and from the Brigham and Women's Hospital anonymized tumor bank (unselected U.S. samples) under Institutional Review Board approved studies. Information on gender, age, and histology was available for most samples. Patient samples were also obtained from patients treated on an open-label clinical trial of gefitinib at Dana-Farber Cancer Institute (13). Responses to gefitinib were defined using standard criteria (See, e.g., A. B. Miller, B. Hoogstraten, M. Staquet, A. Winkler, 1981 Cancer 47, 207-14). IRB approval was obtained for these studies.

[00293] Of the gefitinib-responsive patients, there were two patients who had been previously treated with at least one cycle of chemotherapy, one patient previously treated with radiation therapy, one patient concurrently treated with chemotherapy, and one patient who received no other treatment. For gefitinib-insensitive patients, treatment failure was defined as the appearance of new tumor lesions or the growth of existing tumor lesions in a CT scan after 2 months of gefitinib treatment compared to a baseline CT scan.

[00294] Example 3D: cDNA sequencing of patient samples

[00295] Total RNA is isolated from tissue samples using TrizolTM (Invitrogen, Carlsbad, CA) and is purified using an RNeasyTM mini-elute cleanup kit (Qiagen, Valencia, CA). cDNA is transcribed from 2μg of total RNA with Superscript II Reverse Transcriptase (Invitrogen Life technologies, Carlsbad, CA), according to the manufacturer's recommendations. The cDNA is used as template for subsequent PCR amplifications of EGFR.

[00296] The components of the PCR are: 20mM Tris-HCl(pH 8.4), 50mM KCl, 1.5mM MgCl2, 0.1mM each of dATP, dCTP, dGTP, dTTP, 0.2μM of each primer, and 0.05 units/μl Taq polymerase (Taq Platinum, GIBCO BRL, Gaithersburg, MD). Amplification of fragment "a" requires addition of 4% DMSO to the reaction. The primer sequences are listed in Table S4. Forward and reverse primers are synthesized with 18 base pairs of an overhanging M13 forward and reverse sequences respectively. The thermocycling conditions are: 94°C, 4min; followed by 11 cycles, with denaturing step at 94°C for 20", extension step at 72°C for 20", and with a 20" annealing step that decreased 1°C/ cycle, from 60°C at cycle one to 50°C at cycle 11;

cycle 11 was then repeated 25 times. A 6 minute incubation at 72°C followed by a 4°C soak completes the program.

[00297] An aliquot of the PCR reaction is diluted 1:50 with water. The diluted PCR product is sequenced using an M13 Forward Big Dye Primer kit (Perkin-Elmer/ Applied Biosystems, Foster City, CA), according to the manufacturer's recommendations. The sequencing products are separated on a fluorescent sequencer (model 3100 from Applied Biosystems, Foster City, CA). Base calls are made by the instrument software, and reviewed by visual inspection. Each sequence is compared to the corresponding normal sequence using Sequencher 4.1 software (Gene Codes Corp.).

[00298] Example 3E: Tumor types expressing mutant EGFR

[00299] Two additional mutations in EGFR were found in two different tumor types. An EGFR mutation that substitutes a glycine (G) for a valine (V) at position 857 ("G857V") was identified in Acute Myelogenous Leukemia (AML). An EGFR mutation that substitutes a leucine (L) with a serine (S) at position 883 ("L883S") in a metastatic sarcoma.

[00300] Example 3F: Cell lines

[00301] The effects of gefitinib on NSCLC cell lines in vitro were examined. One cell line, H3255, was particularly sensitive to gefitinib, with an IC50 of 40 nM. Other cell lines had much higher IC50s. For example, a wild type cell line H1666 has an IC50 of 2 uM, which is 50 fold higher than for the mutant cell line When the EGFR from this cell line was sequenced, it contained the L858R missense mutation, while the other cell lines were wild type for EGFR. Much lower concentrations of gefitinib were required to turn off EGFR and also AKT and ERK phosphorylation by EGFR as compared to EGFR wild type cells, which required at least 100 times higher concentrations of gefitinib to achieve the same effect. These findings suggest that the mutant receptor is more sensitive to the effects of gefitinib. Also note here,

[00302] Example 3G: Combination therapies

[00303] Tumor specimens were analyzed from patients with advanced NSCLC treated on the randomized trial of carboplatin/paclitaxel with or without

erlotinib. The clinical portion of this trial demonstrated equivalent survival in the two treatment arms. Tumor specimens were available for sequencing from 228 of the 1076 patients. The preliminary clinical characteristics of these patients is not different from the group as a whole with respect to baseline demographics, response rate, median and overall survival.

[00304] Exons 18-21 of the tyrosine kinase domain were sequenced and 29 mutations, for a mutation frequency of 12.7 percent, were identified.

[00305] As a whole the patients with EGFR mutations have a better survival regardless of whether they received treatment with chemotherapy alone or in combination with erlotinib. These differences are statistically significant with a p value of less than 0.001.These findings raise the possibility the EGFR mutations, in addition to being predictors of response to gefitinib and erlotinib, may also be prognostic for an improved survival.

[00306] (*) Note that the frequency of EGFR mutation in the unselected US patients, 1 of 61, appears to be low when compared to the frequency of reported gefitinib response at 10.4%. This difference has a modest statistical significance ($p = 0.025$ by the chi-squared test). Thus this result could still be due to chance, could be due to a fraction of responders who do not have EGFR mutations, or could be due to failure to detect EGFR mutations experimentally in this tumor collection. If the frequency of EGFR mutation in gefitinib-responsive US patients (5/5) is compared to the expected frequency of gefitinib response (10.4%), the chi-squared probability is again less than 10-12.

EXAMPLE 4

[00307] Study Design:

[00308] We performed a retrospective cohort study of NSCLC patients referred for somatic *EGFR* kinase domain sequencing from August 2004 to January 2005 at Massachusetts General Hospital (MGH), Dana-Farber Cancer Institute (DFCI), and Brigham and Women's Hospital (BWH). These three institutions comprise Dana-Farber/Partners CancerCare (DF/PCC), an academic joint venture cancer center that cares for approximately 1,200 lung cancer patients per year. In

August 2004, *EGFR* kinase domain sequencing was made available for clinical use at DF/PCC. Clinicians could select which patients to refer for testing, however patients needed to have sufficient and appropriate tumor specimens available. Tumor cells had to comprise at least 50% of the specimen based on histologic examination by MGH and BWH reference pathologists, and the specimen had to be from a resection, bronchoscopic biopsy, or core needle biopsy of a primary or metastatic tumor, or a cellblock from pleural fluid. In rare cases, fine needle aspirate samples were determined adequate. Samples could be either paraffin-embedded or frozen tissue. Due to the low incidence of *EGFR* mutations in squamous cell tumors (62) patients with this diagnosis were not eligible for testing.

[00309] We identified patients undergoing *EGFR* testing using the EGFR case log maintained at the Laboratory for Molecular Medicine (LMM), of the Harvard Medical School/Partners HealthCare Center for Genetics and Genomics (CLIA# 22D1005307), the diagnostic testing facility where all sequencing was performed and interpreted. We included all patients referred for *EGFR* testing from DF/PCC with a diagnosis of NSCLC during the study period.

[00310] Patient age, gender, and race were collected from the electronic medical record system. Smoking status, cancer history, *EGFR* kinase domain sequencing results, and subsequent EGFR-TKI treatment plans were documented using structured physician chart review. Specifically, the smoking status and cancer history were obtained from physician and nursing notes. Former smokers were defined as patients who had quit smoking at least one year before their diagnosis of lung cancer and never-smokers were defined as patients who had smoked less than 100 cigarettes in their lifetime. Smokers who had quit within a year of their diagnosis or who were smoking at the time of diagnosis were classified as current smokers. Pack-years of smoking were calculated by multiplying the number of packs smoked per day by the number of years of smoking. Tumor histology and *EGFR* kinase domain sequencing results were obtained from pathology reports. All pathology specimens were centrally reviewed at either MGH or BWH and histology was categorized using the World Health Organization (WHO) classification system (63). Subsequent treatment plans were obtained from physician notes.

[00311] Complete data were available for age, gender, tumor histology, and *EGFR* mutation status. There were missing data for race (12%), tumor stage at time of testing (4%), smoking status (6%), prior treatments (5%), and subsequent EGFR-TKI treatment plans (11%). This study was approved by the Institutional Review Board at DF/PCC.

[00312] EGFR Gene Sequencing:

[00313] Serial sections of either frozen or formalin-fixed, paraffin-embedded (FFPE) tumor tissue were cut and placed on a glass slide. A region of tumor tissue consisting of at least 50% viable tumor cells was identified by a pathologist. FFPE samples were extracted with xylene and ethanol to remove paraffin. Both FFPE and frozen tissue samples were digested with proteinase K overnight. Genomic deoxyribonucleic acid (DNA) was extracted from tissue and peripheral whole blood using standard procedures. Genomic DNA was extracted from saliva samples using the DNA Genotek-Oragene™ saliva kit.

[00314] The kinase domain of *EGFR* (exons 18-24 and flanking intronic regions) was amplified in a set of individual nested polymerase chain reaction (PCR) reactions. The primers used in the nested PCR amplifications are described in Table S1A and B and SEQ ID 1-424 with the addition of universal sequences to the 5' ends of the primers (5' tgtaaaacgacggccagt) (SEQ ID NO. 645). The PCR products were directly sequenced bi-directionally by dye-terminator sequencing. PCR was performed in a 384-well plate in a volume of 15 µl containing 5 ng genomic DNA, 2 mM $MgCl_2$, 0.75 µl DMSO, 1 M Betaine, 0.2 mM dNTPs, 20 pmol primers, 0.2 µl AmpliTaq Gold® (Applied Biosystems), 1X buffer (supplied with AmpliTaq Gold). Thermal cycling conditions were as follows: 95°C for 10 minutes; 95°C for 30 seconds, 60°C for 30 seconds, 72°C for 1 minute for 30 cycles; and 72°C for 10 minutes. PCR products were purified with Ampure® Magnetic Beads (Agencourt).

[00315] Sequencing products were purified using Cleanseq™ Magnetic Beads (Agencourt) and separated by capillary electrophoresis on an ABI 3730 DNA Analyzer (Applied Biosystems). Sequence analysis was performed by Mutation Surveyor (SoftGenetics, State College, PA) and manually by two reviewers. Non-synonymous DNA sequence variants were confirmed by analysis of 3-5 independent

PCR reactions of the original genomic DNA sample. Blood or saliva samples from individuals with non-synonymous DNA sequence variants were analyzed to determine whether the sequence changes were unique to tumor tissue.

[00316] Statistical Analysis:

[00317] We constructed logistic regression models to assess the univariate association between patient demographic and clinical characteristics and *EGFR* mutation status. To identify significant predictors of mutation positive status, we constructed a multivariable logistic regression model including independent variables identified in prior studies as predictive of mutations, specifically gender, race, histology, and smoking status. Six patients were excluded from these analyses due to missing *EGFR* mutation data as a result of PCR failure. All analyses were performed using SAS statistical software (version 8.02, SAS Institute, Cary, NC).

[00318] RESULTS:

[00319] Patient Characteristics:

[00320] Among the 100 patients with NSCLC referred for somatic *EGFR* kinase domain sequencing as part of clinical cancer care during the study period, the mean age was 60.7 years and 63% were female (Table 4). The majority of patients were white (76%) or Asian (7%), and had metastatic disease at the time the test was ordered (67%). Nearly all patients (94%) tested for *EGFR* mutations had adenocarcinoma, adenocarcinoma with bronchioloalveolar carcinoma (BAC) features, or pure BAC. Approximately one third of the patients were never-smokers. Therapy administered prior to the referral for *EGFR* testing included surgery (50%), chest radiotherapy (22%), chemotherapy (47%), and EGFR directed targeted therapy (11%).

[00321] Mutations Identified:

[00322] The average length of time from referral for testing to result availability was 12 business days. The majority of specimens submitted were paraffin-embedded (74%). Six of the 74 (8%) paraffin-embedded specimens failed PCR amplification, while all of the 26 frozen specimens were successfully amplified. Among the 94 patients with interpretable results, 23 (24%) were found to have at least one mutation in the *EGFR* kinase domain, with two of these patients demonstrating

two point mutations each, for a total of 25 mutations identified (Table 5). Among the 23 patients with mutations, 9 (39%) had one or more point mutations, 12 (52%) had in-frame overlapping deletions in exon 19 and two patients (9%) had duplications in exon 20. The point mutations were in exons 18 and 21, and included five 2573T>G (L858R), and one each of 2126A>T (E709V), 2155G>A (G719S), 2156G>C (G719A), 2327G>A (R776H), 2543C>T (P848L), and 2582T>A (L861Q). One of the point mutations (P848L) was detected in both the tumor specimen and in mononuclear cells obtained from a buccal swab. No mutations were detected in exons 22, 23, or 24.

[00323] Predictors of Mutations:

[00324] In our sample, there was no significant association between *EGFR* mutation status and age (p = 0.61), female gender (p = 0.92), Asian race (p = 0.08), or metastatic disease at the time of referral (p = 0.43, Table 4). None of the 6 patients with non-adenocarcinoma tumor histology were found to have mutations. Among the patients with adenocarcinoma, adenocarcinoma with BAC features and pure BAC, there was no association between BAC/BAC features and *EGFR* mutation status (p = 0.35).

[00325] None of the 17 current smokers were found to have a mutation. Never-smokers were significantly more likely to have an *EGFR* mutation than former smokers (odds ratio [OR] = 3.08, 95% confidence interval [CI] 1.09-8.76). The mean number of pack-years smoked was significantly lower among *EGFR* mutation-positive patients (0.7 pack-years) compared to *EGFR* mutation-negative patients (25.0 pack-years, p < 0.001). For each additional pack-year smoked, there was a 4% decrease in the likelihood of having a mutation (OR = 0.96, 95% CI 0.93-0.99).

[00326] The number of pack-years of smoking remained a significant predictor of mutation status after controlling for gender, race, and tumor histology (OR = 0.96, 95% CI 0.93-0.99).

[00327] Subsequent Use of Test Information:

[00328] *EGFR* mutation-positive patients were significantly more likely to have documented plans to receive subsequent EGFR-TKI treatment (86%) than *EGFR* mutation-negative patients (11%, p < 0.001). Clinicians documented that the *EGFR* results affected their prioritization of recommended therapies in 38% of cases. These

cases included 14 (61%) of the 23 mutation-positive patients for whom EGFR-TKI therapy was recommended earlier than it would have been had the test been negative, and 24 (34%) of the 71 mutation-negative patients for whom EGFR-TKI therapy was not recommended, or was recommended later than it would have been had the test been positive.

[00329] EGFR mutation status was more likely to change prioritization of treatment options in patients with metastatic disease (54%) than in patients with local or locally advanced disease (19%, $p = 0.003$). Given this finding, we further analyzed the decision-making process in metastatic patients (Figure 10). Among the 31 patients with metastatic disease whose test results affected treatment recommendations, five mutation-positive patients were offered first-line EGFR-TKI treatment and six mutation-positive patients were offered second-line EGFR-TKI treatment in lieu of chemotherapy. Twenty mutation-negative patients were encouraged to defer EGFR-TKI treatment until third-line treatment or beyond based on their negative *EGFR* test results. Among the 26 patients with metastatic disease whose test results did not affect treatment recommendations, two mutation-negative patients received first-line EGFR-TKI treatment despite their negative results, nine patients including four mutation-positive patients received second or third-line EGFR-TKI treatment, and 15 patients including two mutation-positive patients did not receive a recommendation for an EGFR-TKI. Three of the patients with metastatic disease were participating in trials evaluating first-line EGFR-TKI therapy. Nine of the patients with metastatic disease had previously received or were receiving EGFR-TKIs at the time of *EGFR* testing.

[00330] DISCUSSION:

[00331] We studied the first 100 patients with NSCLC to undergo screening for somatic *EGFR* mutations as part of clinical cancer care at our institution and found that testing was feasible and significantly impacted the treatment of NSCLC patients. Patients harboring *EGFR* mutations were significantly more likely to receive recommendations for EGFR-TKI therapy than patients without mutations. Physicians adjusted their treatment recommendations based on the test results in over one-third of the cases, and were more likely to do so in patients with metastatic disease. In our patient sample, physicians used positive *EGFR* test results to help make the decision

to prioritize EGFR-TKIs over chemotherapy for some patients, especially for first or second-line treatment. However, negative *EGFR* test results did not prevent physicians from administering EGFR-TKIs to selected patients. Many of the patients in whom the test result did not impact clinical decision-making had early stage, resected disease or were already receiving an EGFR-TKI for metastatic disease at the time of testing. This is reasonable since the utility of EGFR-TKIs as adjuvant therapy is not known and there is a benefit to EGFR-TKI therapy in a small number of patients without an identified *EGFR* mutation (65, 66-70, 71).

[00332] Our study also provides evidence that molecular diagnostics can enhance the clinical ability to identify patients with *EGFR* mutations. Many oncologists currently use the clinical characteristics associated with *EGFR* mutations and response to EGFR-TKIs to guide the decision-making process for patients with NSCLC. Indeed, our population of patients referred for *EGFR* testing demonstrated an increased prevalence of such characteristics. For example, 95% of referred patients had adenocarcinoma or BAC tumor histology, compared to 45% in the general NSCLC population (72). While never-smokers comprised 29% of our population, the incidence of never-smokers in the general NSCLC population has been reported as 2-10%, and may be as high as 27% in women with NSCLC (73-75). Similarly, our population consisted of only 17% current smokers, compared to the 38-75% rate of current smoking among newly diagnosed NSCLC patients (75, 78-80). Our clinically selected population consequently had an *EGFR* mutation rate of 24%, which is substantially higher than rates documented by our and other U.S. groups that tested unselected available NSCLC tumor samples (65-66, 81). However, it is important to note that while clinicians appeared to be attempting to select patients for testing that had the clinical characteristics predictive of *EGFR* mutations, the mutation frequency was still only 24%, highlighting the fact that molecular diagnostics increase the information available to make clinical decisions.

[00333] Smoking status was the strongest predictor of *EGFR* mutation status in our patients, with an increase in smoking history associated with a significantly decreased likelihood of harboring an *EGFR* mutation, after controlling for previously described predictors of mutation status. Our results are consistent with other case

series documenting the importance of smoking status in the likelihood of *EGFR* mutations (66, 69, 70, 81, 82). Just as the extremely low prevalence of *EGFR* mutations in squamous cell tumors (62) has shifted testing efforts towards' adenocarcinoma tumors, it may be appropriate to focus future efforts on patients with a low or absent smoking history. However, reports of *EGFR* mutations in patients without typical clinical characteristics advise against strict testing limitations (83). When examining the other clinical characteristics thought to be associated with mutations, we found Asian race and BAC tumor histology to have non-significant trends towards predicting *EGFR* mutation status. The lack of statistical significance in these associations may be due to small sample size.

[00334] The test was feasible and fit into the time constraints of clinical cancer care. Nearly all of the tumors submitted for analysis produced interpretable results. The six specimens that failed PCR amplification were all paraffin-embedded, while none of the frozen specimens failed PCR amplification. When available, fresh frozen tissue is the preferable substrate for *EGFR* mutation testing.

[00335] There have been close to 2,500 NSCLC samples reported thus far that have undergone partial or complete *EGFR* sequence analysis. Our patients demonstrated mutations similar to previous reports, with overlapping exon 19 deletions of 9-23 base pairs and point mutations leading to single amino acid substitutions in exons 18 and 21. Five of the point mutations we found have been described above (E709V, G719S, G719A, L858R, and L861Q). One of the point mutations we found causes an amino acid substitution at a codon where a different amino acid substitution has been previously described (R776H). The E709V and R776H variants were each found in combination with a known gefitinib-sensitizing mutation involving codon 719. The P848L mutation in exon 21 was found in both the somatic and buccal samples, suggesting it may be a germline variant of uncertain significance. The patient was a never-smoking female with adenocarcinoma who had stable disease for 15 months on gefitinib treatment, prior to the *EGFR* mutation testing. When the P848L mutation was revealed, she had recently been found to have progressive disease and was started on erlotinib therapy. No information about response to erlotinib is available at this time.

[00336] The (2253_2276 del) deletion overlaps previously described exon 19 deletions. The deletions in our patients can be categorized into one of two groups: those spanning codons 747-749 at a minimum (amino acid sequence LRE), and those spanning codons 752-759 (Figure 11). Analysis of all exon 19 deletions reported to date suggests that a wide variety of amino acids can be deleted from the TK region spanning codons 747-759. There does not appear to be a required common codon deleted; however, all of the deletions we detected maintained a lysine residue at position 745.

[00337] One of our two exon 20 mutations are in a never-smoking female with recurrent adenocarcinoma who was treated with erlotinib after *EGFR* testing was performed and has had stable disease for two months at this time. The other is a former-smoking male with metastatic adenocarcinoma who was treated with an EGFR-TKI, but could not tolerate it due to severe rash. The identification of clinically relevant *EGFR* mutations in exon 20 underscores the importance of comprehensive sequencing of the TK region of *EGFR*.

[00338] In conclusion, this study demonstrates the feasibility and utility of comprehensive screening of the TK region of the *EGFR* gene for somatic mutations in NSCLC patients as part of clinical cancer care. The result of the test provides useful information regarding clinical predictors of EGFR-TKI response. Current smokers are less likely to harbor a mutation, as are former smokers with a high number of pack-years of smoking history.

EXAMPLE 5

[00339] EGFR GENE TEST FOR NON-SMALL CELL LUNG CANCER, a Standard Operating Prodedure.

[00340] Clinical Indications:

[00341] This test is indicated for individuals with Non-Small Cell Lung Cancer.

[00342] Analytical Principle

[00343] The *EGFR* Gene Test is a genetic test that detects mutations in the kinase domain of *EGFR*. DNA is first obtained from a tumor biopsy. The DNA sequence of 7 exons (18, 19, 20, 21, 22, 23, 24) of *EGFR* is then determined by direct

bi-directional gene sequencing. The sequence obtained is then compared to known *EGFR* sequence to identify DNA sequence changes. If a DNA sequence change is detected in tumor tissue, the test will be repeated on the original tissue sample. If the change has not previously been reported in a gefitinib- or erlotinib- responder, the test will also be conducted with a sample of the individual's blood to determine whether the mutation is constitutive (and therefore likely a normally occurring polymorphism) or occurred somatically in the tumor tissue.

[00344] Specimen Requirements:

[00345] A minimum of 100 ng of DNA is required from tissue sample. Note: Extremely small quantities of DNA may be extracted from tissue samples. The concentration of this DNA may not be accurately quantitated.

[00346] Quality Control:

[00347] Controls used

[00348] Two negative controls (water) and a positive control (human DNA) for each exon are included in the PCR reactions. The negative control should proceed through the entire procedure to ensure that the sequence obtained is not the result of contamination. A pGEM positive control and an ABI array control are included in the sequencing step.

[00349] Control Preparation and Storage:

[00350] The positive control for PCR is either Clontech human DNA or human DNA from an anonymous blood sample and is stored at 4°C. The negative control for the PCR reaction is HyPure Molecular Biology Grade water stored at room temperature. The pGEM positive sequencing reaction control and the ABI array control are stored at –20°C.

[00351] Tolerance Limits and Steps to Take if Individual Control Fails:

[00352] If the positive PCR control fails but the negative controls and samples pass, the PCR results will be designated as pass and sequencing will be performed. If a negative control shows evidence of DNA amplification, the whole reaction will be repeated with a new aliquot of patient's DNA. If the pGEM control fails and the test reactions fail, the sequencing run will be repeated with a second aliquot of the PCR product. If the sequencing controls fail but the test reactions pass,

the sequencing does not need to be repeated. NOTE: Due to the low yield of DNA extraction from paraffin embedded tissue samples, external PCR reactions often do not yield visible products. Internal PCR reactions should yield visible products. The size of the product detected on the gel should be compared to the anticipated sizes (see below) to ensure that the appropriate PCR product has been obtained. If an internal PCR product is not visible on the gel, exon-specific PCR failures should be repeated.

[00353] If PCR amplification for an individual sample fails, a new round of PCR should be attempted with a two-fold increase in input DNA template. If PCR amplification fails again, a new DNA sample for that patient should be acquired if available. If the sample was a paraffin-embedded tissue sample, additional slides should be scraped. If available, more slides than used to generate the original sample should be scraped and digestion in Proteinase K should be allowed to occur for three nights.

[00354] Equipment and Reagents (All reagents stable for one year unless otherwise noted.)

[00355] PCR and Sequencing (in general, PCR and sequencing equipment and reagents are known to those of skill in the art and may be used herein, also noted above).

[00356] Primers: (see Table 6 and 7 below)

Table 6: External PCR Primers:

Exon	Forward Primer Sequence, (5'→3')	SEQ ID NOS	Reverse Primer Sequence, (5'→3')	SEQ ID NOS
18	TCAGAGCCTGTGTTTCTACCAA	653	TGGTCTCACAGGACCACTGATT	646
19	AAATAATCAGTGTGATTCGTGGAG	654	GAGGCCAGTGCTGTCTCTAAGG	647
20	ACTTCACAGCCCTGCGTAAAC	655	ATGGGACAGGCACTGATTTGT	648
21	GCAGCGGGTTACATCTTCTTTC	656	CAGCTCTGGCTCACACTACCAG	649
22	CCTGAACTCCGTCAGACTGAAA	657	GCAGCTGGACTCGATTTCCT	650
23	CCTTACAGCAATCCTGTGAAACA	658	TGCCCAATGAGTCAAGAAGTGT	651
24	ATGTACAGTGCTGGCATGGTCT	659	CACTCACGGATGCTGCTTAGTT	652

TABLE 7: Internal PCR Primers:

Exon	Forward Primer Sequence, (5'→3')	Reverse Primer Sequence, (5'→3')	Product Length (bp)
18	TCCAAATGAGCTGGCAAGTG (SEQ ID NO 660)	TCCCAAACACTCAGTGAAACAAA (SEQ ID NO 667)	397
19	GTGCATCGCTGGTAACATCC (SEQ ID NO 661)	TGTGGAGATGAGCAGGGTCT (SEQ ID NO 668)	297
20	ATCGCATTCATGCGTCTTCA (SEQ ID NO 662)	ATCCCCATGGCAAACTCTTG (SEQ ID NO 669)	378
21	GCTCAGAGCCTGGCATGAA (SEQ ID NO 663)	CATCCTCCCCTGCATGTGT (SEQ ID NO 670)	348
22	TGGCTCGTCTGTGTGTGTCA (SEQ ID NO 664)	CGAAAGAAAATACTTGCATGTCAGA (SEQ ID NO 671)	287
23	TGAAGCAAATTGCCCAAGAC (SEQ ID NO 665)	TGACATTTCTCCAGGGATGC (SEQ ID NO 672)	383
24	AAGTGTCGCATCACCAATGC (SEQ ID NO 666)	ATGCGATCTGGGACACAGG (SEQ ID NO 673)	302
F primer linker	tgtaaaacgacggccagt (SEQ ID NO 645)	5' end of all forward primers	18
R primer linker	aacagctatgaccatg (SEQ ID NO 674)	5' end of all reverse primers	16

[00357] Precautions

Table 8

Task	Instruction(s)	Risk
1. PCR Setup	Use PCR Hood Use dedicated pipets and filtered tips Only open reagents in the hood	Contamination of PCR reaction
2. Use of PCR Hood	Do not use any post-PCR samples or reagents in the hood	Contamination of PCR reaction

[00358] Preparing PCR Reaction Mix for External PCR

[00359] All procedures performed in PCR hood for genomic DNA, not the clean hood.

1. Thaw out Taq Gold and dNTP on ice.

2. Prepare the master mix in a tube (eppendorf or 15mL tubes) using the table below. Water, Betaine, 10X Buffer, $MgCl_2$, DMSO, Taq Gold and dNTP should be added in the order listed. It is very important to mix the reagents by pipetting up-and-down gently while adding each reagent.
3. DNA should be added to the master mix before aliquoting. After making the large volume of master mix, aliquot 96 ul (enough for 8 rxns) to a separate tube for each patient or control. Add 8 ul of DNA at 5 ng/ul to the 96 ul of mastermix. 13ul can then be added to the individual wells of the plate or put in strip tubes and pipetted with a multi-channel pipettor.
4. For a full 384-well plate of reactions, make enough master mix for about 415 reactions.
5. Spin the plate of master mix to get rid of air bubbles.
6. If using a large set of primers, it would help to have them in 96-well plates with forward primers and reverse primers in separate plates.
7. Add the primers using a multi-channel pipette. Make sure to mix by pipetting up-and-down gently.
8. Spin the plate to get rid of any air bubbles.
9. Use the cycle below to amplify.

Note: PCR is done in 384-well plates.

TABLE 9

Reagent	*Volume per reaction (μL)*
Autoclaved ddH_2O	4.90
5M Betaine	3.00
10X Buffer	1.50
Magnesium Chloride	1.50
DMSO	0.75
Taq	0.20
dNTP	0.15
PCR Forward Primer1 (conc. 20pmol/uL)	1.00
PCR Reverse Primer2 (conc. 20pmol/uL)	1.00
DNA (conc. 5ng/uL)	1.00
Total volume of PCR reaction	15.00

TABLE 10: PCR Amplification Cycle

Activate Taq Gold	10 minutes	95 °C	
Denature	30 seconds	95 °C	30 cycles
Anneal	30 seconds	60 °C	
	1 minutes	72 °C	
Extend	10 minutes	72 °C	
Hold	∞	4 °C	

Note: A cleanup is not necessary after performing the external PCR.

[00360] Preparing PCR Reaction Mix for Internal PCR

[00361] The internal PCR set up is almost the same as the external PCR with a few exceptions.

1. Make the large volume of master mix as described for external PCR in the PCR hood.
2. Aliquot MM to 7 strip tubes and multichannel pipette 12ul into the 384-well plate.
3. Add 1ul each of forward and reverse internal primers. Temporarily seal plate.
4. Remove from hood, spin down plate and proceed to post PCR set-up area.
5. Use dedicated pipettes to aliquot 1ul of external PCR product into each reaction.
6. Heat seal and spin again.
7. Run same amplification cycle as external.

[00362] Run PCR products on a 1% gel before clean-up. Determine Pass/Failed exons for repeat PCR.

[00363] Clean-up Internal PCR Using Ampure Magnetic Bead Clean-up

[00364] Cleanup

1. Vortex the plate of Ampure magnetic beads till there is no deposit of beads.
2. It is very important that the temperature of the Ampure beads is at room temperature.

3. Use the 384-well Ampure protocol on the Biomek and change the volume of reaction to 12uL to accommodate reagents used for cleanup. If this is not done, an error will be generated.

4. After the program is complete, hydrate plate with 20 uL of autoclaved ddH_2O per well. While adding water, make sure to mix by pipetting up-and-down gently.

5. Spin the plate to get rid of any air bubbles.

6. Place the plate on a magnet to separate out the beads. Now you should be able to take up 1 uL of the DNA to setup sequencing reactions. Save the rest at -20°C for future use.

[00365] Sequencing Protocol

[00366] Preparing Sequencing Reaction Mix

1. Thaw out BigDye 3.1 in a dark place, on ice.
2. Prepare the master mix in a tube (eppendorf or 15mL tubes) using the table below. Water, buffer, DMSO and BigDye should be added in the order listed.
3. It is very important to mix the reagents by pipetting up-and-down gently while adding each reagent.
4. When using a universal primer for sequencing, the primer can also be added to the master mix at this time. If the primer is unique it should be added individually after the master mix is in the 384-well plate.
5. Usually for a full 384-well plate of reactions, make enough master mix for about 415 reactions.
6. Once the master mix is setup divide the mix into 8 wells of strip tubes. (Do not use reservoirs to aliquot master mix. That would be a waste of reagents.) 7.
7. Now a multi-channel pipette can be used to aliquot the master mix into the 384-well plate
8. Spin the plate of master mix to get rid of air bubbles.
9. Add the PCR product to be sequenced, using a multi-channel pipette. Make sure to mix by pipetting up-and-down.
10. Spin the plate to get rid of any air bubbles.

11. Use the cycle below to amplify.

TABLE 11

Reagent	*Volume per reaction (μL)*
Autoclaved ddH_2O	4.38
5X ABI Buffer	3.65
DMSO	0.50
ABI BigDye 3.1	0.35
Sequencing Primer concentration	0.12
DNA from Internal PCR reaction	1.00
Total Volume of reaction	10.00

TABLE 12: Amplification Cycle for Sequencing

Denature	10 seconds	96 °C	25 cycles
Anneal	5 seconds	50 °C	
Extend	4 minutes	60 °C	
Hold	∞	4 °C	

[00367] Clean-up via Cleanseq Magnetic Bead Clean-up

1. Vortex the plate of Cleanseq magnetic beads till there is no deposit of beads.
2. Use the Cleanseq 384-well plate program on the Biomek to clean-up the samples.
3. Once the program is done, save the original plate at -20 °C. The new plate with the clean samples is ready to go on the ABI 3730.

[00368] (Note: If the PCR products are shorter than 300 bps you might have to dilute the sample before putting it on the 3730)

[00369] Create Mutation Surveyor templates for the EGFR test and save them on LMM/Sequencing/ Sequences-MS Review/EGFR.

[00370] Repeat Result Criteria

[00371] All positive results are repeated by amplifying and sequencing the specific exon(s) in which a DNA sequence change has been detected from a second aliquot of patient DNA derived from the original tissue sample. In addition, DNA extracted from a sample of the patient's blood should be run in parallel to compare with tumor tissue if the sequence change detected has not previously been detected in a gefitinib- or erlotinib-responder.

[00372] Any exon that did not produce clear sequence will be repeated either from extraction, PCR or sequencing, based on the specific technical problems.

[00373] Assay Parameters

[00374] Sensitivity of the Test – Somatic *EGFR* kinase domain mutations have been found in approximately 13% of individuals with NSCLC (Paez JG et al., 2004). In addition, somatic *EGFR* kinase domain mutations have been found in 13/14 (92.8%) individuals with NSCLC that were gefitinib-responsive (Paez JG et al., 2004, Lynch, et al., 2004). Validation of the technical sensitivity of the test demonstrated 100% sensitivity to known mutations and validation of the sequencing platform in our lab shows 100% sensitivity (see "Accuracy of the Technique" below). The sensitivity for mutation detection of mosaic samples has been determined to be 25% (ie, heterozygous mutations can be detected when present at 50% of a cell mixture). We have found that up to 20% of paraffin-embedded tissue do not yield high quality DNA. We are unable to obtain sequence information from these samples.

[00375] Specificity of the Test – To date, published literature indicates that no individuals with a somatic mutation in EGFR were not responsive to gefitinib (11/11). The chance of finding a mutation due to an artifact of bi-directional sequencing is close to 0% (see "Accuracy of the Technique" below). As such, the specificity of the test is approximately 100%.

[00376] Accuracy of the Technique – The technique of DNA sequencing is the gold standard in molecular diagnostics. This lab uses the ABI 3730 DNA Analyzer that has a reported accuracy of 98.5%. Combining this with bi-directional sequencing, automated chromatogram analysis with Mutation Surveyor, and manual analysis of false positives, we have achieved an accuracy rate of 100%. This is based upon an analysis of over 100,000 bases of raw sequence. For details of this assessment, see our Quality Assurance Program manual.

[00377] Note: We do not assume that these results guarantee 100% accuracy of this platform. It is known that sequencing errors can occur and, as such, we report our accuracy to be 99.99% that has been found by large scale sequencing projects (Hill et al. 2000).

[00378] Reproducibility of the Test – Due to the accuracy of the test, when results are achieved, they are reproducible equal to the accuracy of the test (99.99%). However, on occasion, the test can fail due to factors listed below (see Limitations of Method) or because of PCR or sequencing failure due to unexplained technical reasons. In these cases, no results are achieved and the assay is repeated until a result is achieved or the patient specimen is deemed unacceptable. Specific rates of failure of each assay step and of specimens can be found in the validation reports in our Quality Assurance Program manual.

[00379] Normal Range of the Results – The normal sequence of the *EGFR* gene can be found online using GenBank accessions: NT_033968.5 (genomic sequence) and NM_005228.3 (mRNA sequence).

[00380] Limitations of Method:

[00381] Large deletions spanning one or more exons will not be detected by the sequencing method, particularly if present in heterozygosity. Mutations in the *EGFR* gene outside of the kinase domain will not be detected by this assay. Inhibitors may be present in the DNA sample preventing amplification by PCR. Degraded DNA may not produce analyzable data and re-submission of the specimen may be required. Rare sequence variations or secondary structures of the targeted primer sequences could affect PCR amplification and therefore mutation(s) could be missed in that region of one allele.

EXAMPLE 6

[00382] Gefitinib (Iressa) is a tyrosine kinase inhibitor that targets the epidermal growth factor receptor (EGFR), and induces dramatic clinical responses in non-small cell lung cancers (NSCLCs) with activating mutations within the EGFR kinase domain. We report that these mutant EGFRs selectively activate Akt and STAT signaling pathways, which promote cell survival, but have no effect on Erk/MAPK signaling, which induces proliferation. NSCLCs expressing mutant EGFRs underwent extensive apoptosis following siRNA-mediated knockdown of the mutant EGFR or treatment with pharmacological inhibitors of Akt and STAT signaling, and were relatively resistant to apoptosis induced by conventional chemotherapeutic drugs. Thus, mutant EGFRs selectively transduce survival signals on which NSCLCs become dependent; consequently, inhibition of those signals by Gefitinib may underlie striking clinical responses.

[00383] Receptor tyrosine kinases of the EGFR family regulate essential cellular functions including proliferation, survival, migration, and differentiation, and appear to play a central role in the etiology and progression of solid tumors (R. N. Jorissen et al., *Exp. Cell Res.* **284**, 31 (2003), H. S. Earp, T. L. Dawson, X. Li, H. Yu, *Breast Cancer Res. Treat.* **35**, 115 (1995)). EGFR is frequently overexpressed in breast, lung, colon, ovarian, and brain tumors, prompting the development of specific pharmacological inhibitors, such as Gefitinib, which disrupts EGFR kinase activity by binding the ATP pocket within the catalytic domain (A. E. Wakeling et al., *Cancer Res.* **62**, 5749 (2002)). Gefitinib has induced dramatic clinical responses in approximately 10% of patients with chemotherapy-refractory NSCLC (J. Baselga et al., *J. Clin. Oncol.* **20**, 4292 (2002), M. Fukuoka et al., *J. Clin. Oncol.* **21**, 2237 (2003), G. Giaccone et al., *J Clin Oncol.* **22**, 777 (2004), M. G. Kris et al., *JAMA* **290**, 2149 (2003)). Virtually all Gefitinib-responsive lung cancers harbor somatic mutations within the EGFR kinase domain, whereas no mutations have been seen in non-responsive cases (T. J. Lynch et al., *N. Engl. J. Med.* **350**, 2129 (2004), J. G. Paez et al., *Science* **304**, 1497 (2004).) These heterozygous mutations include small in-frame deletions and missense substitutions clustered within the ATP-binding pocket.

[00384] Using transient transfections of mutant EGFRs, we showed previously that both types of mutations lead to increased EGF-dependent receptor activation, as measured by autophosphorylation of Y1068, one of the prominent C-terminal phosphorylation sites of EGFR. (T. J. Lynch et al., *N. Engl. J. Med.* **350**, 2129 (2004).

[00385] To enable studies of qualitative differences in signaling by mutant EGFRs, we generated stable lines of non-transformed mouse mammary epithelial cells (NMuMg) expressing wild-type or mutant EGFRs, and analyzed EGF–mediated autophosphorylation of multiple tyrosine residues linked to activation of distinct downstream effectors (R. N. Jorissen et al., *Exp. Cell Res.* **284**, 31 (2003)). Cell lines were generated that expressed either wild-type EGFR or one of two recurrent mutations detected in tumors from Gefitinib-responsive patients: the missense mutation L858R and the 18bp in-frame deletion, delL747-P753insS. Significantly different tyrosine phosphorylation patterns were observed between wild-type and the two mutant EGFRs at several C-terminal sites. EGF-induced phosphorylation of Y1045 and Y1173 was virtually indistinguishable between wild-type and mutant EGFRs, whereas phosphorylation of Y992 and Y1068 was substantially increased in both mutants. Interestingly, Y845 was highly phosphorylated in the L858R missense mutant, but not in the wild-type or the deletion mutant, and hence appears to be unique in distinguishing between the two types of EGFR mutations. The differential EGF-induced tyrosine phosphorylation pattern seen with wild-type and mutant receptors was reproducible in transiently transfected COS7 cells, ensuring against potential cell type specific effects.

[00386] Thus, Gefitinib-sensitive mutant EGFRs transduce signals that are qualitatively distinct from those mediated by wild-type EGFR. These differences may result directly from structural alterations within the catalytic pocket affecting substrate specificity, or from altered interactions with accessory proteins that modulate EGFR signaling.

[00387] The establishment of cell lines stably transfected with mutant EGFRs made it possible to compare the phosphorylation status of the major downstream targets of EGFR in a shared cellular background. EGF-induced activation

of Erk1 and Erk2, via Ras, of Akt via PLCγ/PI3K, and of STAT3 and STAT5 via JAK2, are essential downstream pathways mediating oncogenic effects of EGFR (R. N. Jorissen et al., *Exp. Cell Res.* **284**, 31 (2003)). EGF-induced Erk activation was essentially indistinguishable among cells expressing wild-type EGFR or either of the two activating EGFR mutants. In contrast, phosphorylation of both Akt and STAT5 was substantially elevated in cells expressing either of the mutant EGFRs. Increased phosphorylation of STAT3 was similarly observed in cells expressing mutant EGFRs. The unaltered Erk activation by the mutant EGFRs is consistent with the absence of increased phosphorylation of Y1173, an important docking site for the Shc and Grb-2 adaptors that leads to Ras activation and subsequent Erk phosphorylation (R. N. Jorissen et al., *Exp. Cell Res.* **284**, 31 (2003)). The increased Akt and STAT phosphorylation following activation of the mutant EGFRs is consistent with the increase in Y992 and Y1068 phosphorylation, both of which have been previously linked to Akt and STAT activation (R. N. Jorissen et al., *Exp. Cell Res.* **284**, 31 (2003)). Thus, the selective EGF-induced autophosphorylation of C-terminal tyrosine residues within EGFR mutants is well correlated with the selective activation of downstream signaling pathways.

[00388] To extend these observations to lung cancer cells in which *EGFR* mutations appear to drive tumorigenesis, we studied lines derived from five NSCL tumors. NCI-H1975 carries the recurrent heterozygous missense mutation L858R and NCI-H1650 has the in-frame deletion delE746-A750, whereas NCI-358, NCI-H1666, and NCI-H1734 express wild-type *EGFR*. As in transfected cells, EGF-induced autophosphorylation of Y992 and Y1068 was markedly elevated in the two lines with endogenous *EGFR* mutations, as was phosphorylation of Akt and STAT5, but not Erk.

[00389] The oncogenic activity of EGFR reflects the activation of signals that promote both cell proliferation and cell survival (S. Grant, L. Qiao, P. Dent, *Front. Biosci.* 7, d376 (2002)). While these pathways exhibit overlap, Ras-mediated activation of the Erk kinases contributes substantially to the proliferative activity of EGFR, whereas activation of Akt and STATs is largely linked to an anti-apoptotic function (S. Grant, L. Qiao, P. Dent, *Front. Biosci.* 7, d376 (2002), F. Chang et al.,

Leukemia **17**, 1263 (2003), F. Chang et al., *Leukemia* **17**, 590 (2003), F. Chang et al., *Int. J. Oncol.* **22**, 469 (2003), V. Calo et al., *J. Cell Physiol.* **197**, 157 (2003), T. J. Ahonen et al., *J. Biol. Chem.* **278**, 27287 (2003)). The two lung cancer cell lines harboring *EGFR* mutations exhibited a proliferative response to EGF at low serum concentrations that was not observed in cells with wild-type receptors. However, their proliferation rate and cell density at confluence were comparable at normal serum concentrations.

SiRNA

[00390] In contrast, apoptotic pathways were markedly different in lung cancer cells with mutant EGFRs: siRNA-mediated specific inactivation of mutant *EGFR* in these cell lines resulted in rapid and massive apoptosis. About 90% of NCI-H1975 cells transfected with L858R-specific siRNA died within 96 hours, as did NCI-H1650 cells transfected with delE746-A750-specific siRNA. SiRNA specific for either *EGFR* mutation had no effect on cells expressing the alternative mutation, and siRNA that targets both wild-type and mutant *EGFR* had minimal effect on the viability of cells expressing only wild-type receptor, but induced rapid cell death in lines expressing *EGFR* mutants. The ability of siRNAs to specifically target the corresponding *EGFR* alleles was confirmed in transfected COS7 cells by immunoblotting. Thus, expression of mutant EGFRs appears essential for suppression of pro-apoptotic signals in lung cancers harboring these mutations. The fact that lung cancer cells expressing only wild-type receptors do not display a similar dependence on EGFR expression may also account for the relative Gefitinib-insensitivity of human tumors that overexpress wild-type EGFR.

[00391] The effectiveness of Gefitinib in lung cancers harboring mutant EGFRs may reflect both its inhibition of critical anti-apoptotic pathways on which these cells have become strictly dependent, as well as altered biochemical properties of the mutant receptors. We previously reported that mutant EGFRs are more sensitive to Gefitinib inhibition of EGF-dependent autophosphorylation than wild-type receptors (T. J. Lynch et al., *N. Engl. J. Med.* **350**, 2129 (2004)). This increased drug sensitivity by mutant receptors was also observed for both Erk and STAT5

activation. Thus, while EGF-induced signaling by mutant receptors demonstrates selective activation of downstream effectors via differential autophosphorylation events, their enhanced inhibition by Gefitinib is uniform, and may reflect altered drug binding to the mutant ATP pocket.

[00392] To establish the relevance of increased Akt and STAT signaling in EGFR-mediated NSCLC survival, we targeted these pathways with specific pharmacological inhibitors. Lung cancer cells harboring *EGFR* mutations were 100-fold more sensitive to Gefitinib than cells with wild-type receptor. Cells expressing mutant EGFRs were also more sensitive to pharmacological inhibition of Akt or STAT signaling than cells expressing only wild-type EGFR. While EGFR-mutant lung cancer cells exhibited increased sensitivity to disruption of Akt/STAT-mediated anti-apoptotic signals, they demonstrated markedly increased resistance to cell death signals induced by the commonly used chemotherapeutic agents doxorubicin and cisplatin, and the pro-apoptotic Fas-ligand.

[00393] Enhanced Akt/STAT signaling in cells with mutant EGFR might therefore provide an additional therapeutic target, while raising the possibility that conventional chemotherapy may be less effective against these tumors.

[00394] "Oncogene addiction" has been proposed to explain the apoptosis of cancer cells following suppression of a proliferative signal on which they have become dependent (I. B. Weinstein, *Science* **297**, 63 (2002)). Interestingly, Imatinib (Gleevec) efficiently triggers cell death in chronic myeloid leukemias expressing the BCR-ABL translocation product and in gastrointestinal stromal tumors expressing activating c-Kit mutations, both of which exhibit frequently constitutive STAT activation that is effectively inhibited by the drug (T. Kindler et al., *Leukemia* **17**, 999 (2003), G. P. Paner et al., *Anticancer Res.* **23**, 2253 (2003)). Similarly, in lung cancer cells with EGFR kinase mutations, Gefitinib-responsiveness may result in large part from its effective inhibition of essential anti-apoptotic signals transduced by the mutant receptor.

Materials and Methods

Immunoblotting

[00395] Lysates from cultured cells were prepared in ice-cold RIPA lysis solution (1% Triton X-100, 0.1% SDS, 50 mM Tris-Hcl, pH 7.4, 150 mM NaCl, 1 mM EDTA, 1 mM EGTA, 10 mM β-glycerol-phosphate, 10 mM NaF, 1 mM Na-orthovanadate, containing protease inhibitors. Debris was removed by centrifugation in a microfuge at 12,000 x g for 10 min at 4°C. Clarified lysates were boiled in gel loading buffer and separated by 10% SDS-PAGE. Proteins were electrotransferred to nitrocellulose and detected with specific antibodies followed by incubation with horseradish peroxidase-conjugated secondary goat antibody (Cell signaling (Beverly, MA; 1:2000) and development with enhanced chemiluminescence (DuPont NEN) followed by autoradiography. The phospho-EGFR Y845, Y992, Y1045, Y1068, phospho-STAT5 (tyr694), phospho-AKT(Ser473), phospho-ERK1/2(Thr202/Tyr204), AKT, STAT5, and ERK1/2 antibodies were obtained from New England Biolabs (Beverly, MA). The total EGFR Ab-20 antibody was obtained from NeoMarkers (Fremont, CA). The phospho-EGFR Y1173 antibody was from Upstate Biotechnology (Lake Placid, NY) and the total phosphotyrosine antibody PY-20 was from Transduction Laboratories (Lexington, KY). All antibodies were used at a 1:1000 dilution.

EGFR expression vectors

[00396] Full-length EGFR expression constructs encoding the wild type, L858 or del L747-P753insS mutations were sub-cloned using standard methods into plasmid pUSEamp. All constructs were confirmed by DNA sequence analysis.

Cell lines and transfections

[00397] COS7 cells and NMuMg (normal mouse mammary epithelial) cells were grown in DMEM (Dulbecco's modified Eagle's media) with 10% fetal calf serum in the presence of 2mM L-glutamine and 50 U/ml penicillin/streptomycin. The NCI-H358, NCI-H1650, NCI-H1734, NCI-H1666, and NCI-H1975 human lung

cancer cell lines were obtained from the American Type Culture Collection collection and were grown in RPMI1640 with 10% fetal bovine serum, 2mM L-glutamine, 50 U/ml penicillin/streptomycin and 1mM sodium pyruvate. They are referred to in the text, in an abbreviated manner, as H358, H1650, H1734, H1666, and H1975, respectively. Transient transfection of COS7 cells was performed using Lipofectamine 2000 (Invitrogen; Carlsbad, CA). Plasmid (1 μg) was transfected into cells at 80% confluence on a 10 cm dish. After 12 hours, the cells were harvested and reseeded in 12-well plates in the absence of serum. The following day, cells were stimulated with 30ng/ml of EGF. Stable NMuMg cell lines were prepared by co-transfecting the EGFR expression constructs with the drug-selectable plasmid pBABE puro, followed by selection in 3 ug/ml puromycin. Pools of drug-resistant cells were used for analysis. Expression of EGFR in stably transfected cells was confirmed by immunoblotting.

SiRNA-mediated "knockdown" of EGFR expression

[00398] SiRNA for EGFR L858R was designed to target the nucleotide sequence CACAGATTTTGGGCGGGCCAA (SEQ ID NO.: 688), while the GCTATCAAAACATCTCCGAAA (SEQ ID NO.: 689) sequence was used for the delE745-A750 (Qiagen; Valencia, CA). To target all forms of EGFR, commercially prepared siRNA corresponding to human wild-type EGFR was obtained from Dharmacon (Lafayette, CO). Transfection of siRNAs was performed with Lipofectamine 2000 (Invitrogen) as per the manufacturer's instructions. Cells were assayed for viability after 96 hours using the MTT assay.

Apoptosis assay

[00399] 10,000 cells were seeded into individual wells of a 96-well plate. After 6 hours, the medium was changed and the cells were maintained in the presence of increasing concentrations of doxorubicin (Sigma; St. Louis, MO), cisplatin (Sigma), Fas-ligand (human activating, clone CH11; Upstate Biotechnology), Ly294002 (Sigma), or AG490 (Calbiochem; La Jolla, CA). After 96 hours, the viability of cells was determined using the MTT assay. For caspase immunostaining,

10,000 cells were seeded onto 10 mm coverslips. The next day they were transfected with siRNA (see previous section for details). After 72 hours the cells were fixed in 4% paraformaldehyde at room temperature for 10 min. They were subsequently permeabilized for 5 min in 0.5% Triton X-100 and blocked for 1hr in 5% normal goat serum (NGS). The coverslips were then incubated overnight at 4°C in primary antibody (cleaved caspase-3 Asp175 5A1 from Cell Signaling) at a 1:100 dilution. The next day the coverslips were washed 3 times in PBS and incubated for 1 hour with secondary antibody (goat anti-rabbit Texas-red conjugated; from Jackson Immunoresearch; West Grove, PA) at a 1:250 dilution in 5% normal goat serum and 0.5 μg/ml of DAPI (4',6-Diamidino-2-phenylindole). After 3 washes in PBS the coverslips were mounted with ProLong Gold anti-fade reagent from Molecular Probes (Eugene, OR).

Cell viability assay

[00400] 10 μl of 5mg/ml MTT (Thiazolyl blue; Sigma) solution was added to each well of a 96-well plate. After 2 hours of incubation at 37°C, the medium was removed and the MTT was solubilized by the addition of 100μl of acidic isopropanol (0.1N HCL) to each well. The absorbance was determined spectophotometrically at 570nm.

Growth curve

[00401] Growth curves for H-358, H-1650, H-1734, and H-1975 cells were obtained by seeding 1000 cells in individual wells of 96-well plates. Each cell line was plated in 8 separate wells. On consecutive days, the cells were fixed in 4% formaldehyde and stained with 0.1 %(w/v) crystal violet solution. The crystal violet was then solubilized in 100μl of 10% acetic acid, and the absorbance was measured at 570nm using a plate reader to determine the relative cell number.

Mutation identification

[00402] To identify sporadic NSCLC cell lines harboring mutations within EGFR, we sequenced exons 19 and 21 within a panel of 15 NSCLC cell-lines, as

described above. Cell lines were selected for analysis based on their derivation from tumors of bronchoalveolar histology irrespective of smoking history (NCI-H358, NCI-H650, NCI-H1650), or from adenocarcinomas arising within non-smokers (NCI-H1435, NCI-H1563, NCI-H1651, NCI-H1734, NCI-H1793, NCI-H1975, NCI-H2291, NCI-H2342, NCI-H2030, NCI-H1838, NCI-H2347, NCI-H2023). NCI-H1666 has been reported to harbor only wild type EGFR (see examples above). All cell lines are available from the American Type Culture Collection.

[00403] The references cited herein and throughout the specification are incorporated herein by reference in their entirety.

REFERENCES

1. Schiller JH, Harrington D, Belani CP, ct al. Comparison of four chemotherapy regimens for advanced non-small cell lung cancer. N Engl J Med 2002; 346:92-98.

2. Druker BJ, Talpaz M, Resta DJ et al. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in Chronic Myeloid Leukemia. N Engl J Med 2001;344:1031-1037.

3. Arteaga CL. ErbB-targeted therapeutic approaches in human cancer. Exp Cell Res. 2003; 284:122-30.

4. Jorissen RN, Walker F, Pouliot N, Garrett TP, Ward CW, Burgess AW. Epidermal growth factor receptor: mechanisms of activation and signaling. Exp Cell Res 2003;284:31-53

5. Luetteke NC, Phillips HK, Qui TH, Copeland NG, Earp HS, Jenkins NA, Lee DC. The mouse waved-2 phenotype results from a point mutation in the EGF receptor tyrosine kinase. Genes Dev 1994;8:399-413.

6. Nicholson RI, Gee JMW, Harper ME. EGFR and cancer prognosis. Eur J Cancer. 2001;37:S9-15

7. Wong AJ, Ruppert JM, Bigner SH, et al. Structural alterations of the epidermal growth factor receptor gene in human gliomas. Proc Natl Acad Sci. 1992;89:2965-2969.

8. Ciesielski MJ, Genstermaker RA. Oncogenic epidermal growth factor receptor mutants with tandem duplication: gene structure and effects on receptor function. Oncogene 2000; 19:810-820.

9. Frederick L, Wang W-Y,Eley G, James CD. Diversity and frequency of epidermal growth factor receptor mutations in human glioblastomas. Cancer Res 2000; 60:1383-1387.

10. Huang H-JS, Nagane M. Klingbeil CK, et al. The enhanced tumorigenic activity of a mutant epidermal growth factor receptor common in human cancers is mediated by threshold levels of constitutive tyrosine phophorylation and unattenuated signaling. J Biol Chem 1997;272:2927-2935

11. Pegram MD, Konecny G, Slamon DJ. The molecular and cellular biology of HER2/neu gene amplification/overexpression and the clinical development of herceptin (trastuzumab) therapy for breast cancer. Cancer Treat Res 2000;103:57-75.

12. Ciardiello F, Tortora G. A novel approach in the treatment of cancer targeting the epidermal growth factor receptor. Clin Cancer Res. 2001;7:2958-2970

13. Wakeling AE, Guy SP, Woodburn JR et al. ZD1839 (Iressa): An orally active inhibitor of Epidermal Growth Factor signaling with potential for cancer therapy. Cancer Res 2002;62:5749-5754.

14. Moulder SL, Yakes FM, Muthuswamy SK, Bianco R, Simpson JF, Arteaga CL. Epidermal growth factor receptor (HER1) tyrosine kinase inhibitor ZD1839 (Iressa) inhibits HER2/neu (erbB2)-overexpressing breast cancer cells *in vitro* and *in vivo*. Cancer Res 2001;61:8887-8895.

15. Moasser MM, Basso A, Averbuch SD, Rosen N. The tyrosine kinase inhibitor ZD1839 ("Iressa") inhibits HER2-driven signaling and suppresses the growth of HER-2 overexpressing tumor cells. Cancer Res 2001;61:7184-7188.

16. Ranson M, Hammond LA, Ferry D, et al. ZD1839, a selective oral epidermal growth factor receptor-tyrosine kinase inhibitor, is well tolerated and active in patients with solid, malignant tumors: results of a phase I trial. J Clin Oncol. 2002; 20: 2240-2250.

17. Herbst RS, Maddox A-M, Rothernberg ML, et al. Selective oral epidermal growth factor receptor tyrosine kinase inhibitor ZD1839 is generally well tolerated and has activity in non-small cell lung cancer and other solid tumors: results of a phase I trial. J Clin Oncol. 2002;20:3815-3825.

18. Baselga J, Rischin JB, Ranson M, et al. Phase I safety, pharmacokinetic and pharmacodynamic trial of ZD1839, a selective oral Epidermal Growth Factor Receptor tyrosine kinase inhibitor, in patients with five selected solid tumor types. J Clin Onc 2002;20:4292-4302.

19. Albanell J, Rojo F, Averbuch S, et al. Pharmacodynamic studies of the epidermal growth factor receptor inhibitor ZD1839 in skin from cancer patients: histopathologic and molecular consequences of receptor inhibition. J Clin Oncol. 2001; 20:110-124.

20. Kris MG, Natale RB, Herbst RS, et al. Efficacy of Gefitinib, an inhibitor of the epidermal growth factor receptor tyrosine kinase, in symptomatic patients with non-small cell lung cancer: A randomized trial. JAMA 2003;290:2149-2158.

21. Fukuoka M, Yano S, Giaccone G, et al. Multi-institutional randomized phase II trial of gefitinib for previously treated patients with advanced non-small-cell lung cancer. J Clin Oncol 2003;21:2237-2246.

22. Giaccone G, Herbst RS, Manegold C, et al. Gefitinib in combination with gemcitabine and cisplatin in advanced non-small-cell lung cancer: A phase III trial-INTACT 1. J Clin Oncol 2004;22:777-784.

23. Herbst RS, Giaccone G, Schiller JH, et al. Gefitinib in combination with paclitaxel and carboplatin in advanced non-small-cell lung cancer: A phase III trial – INTACT 2. J Clin Oncol 2004;22:785-794.

24. Rich JN, Reardon DA, Peery T, et al. Phase II Trial of Gefitinib in recurrent glioblastoma. J Clin Oncol 2004;22:133-142

25. Cohen MH, Williams GA, Sridhara R, et al. United States Food and Drug Administration Drug Approval Summary: Gefitinib (ZD1839;Iressa) Tablets. Clin Cancer Res. 2004;10:1212-1218.

26. Cappuzzo F, Gregorc V, Rossi E, et al. Gefitinib in pretreated non-small-cell lung cancer (NSCLC): Analysis of efficacy and correlation with HER2 and

epidermal growth factor receptor expression in locally advanced or Metastatic NSCLC. J Clin Oncol. 2003;21:2658-2663.

27. Fitch KR, McGowan KA, van Raamsdonk CD, et al. Genetics of Dark Skin in mice. Genes & Dev 2003;17:214-228.

28. Nielsen UB, Cardone MH, Sinskey AJ, MacBeath G, Sorger PK. Profiling receptor tyrosine kinase activation by using Ab microarrays. Proc Natl Acad Sci USA 2003;100:9330-5.

29. Burgess AW, Cho H, Eigenbrot C, et al. An open-and-shut case? Recent insights into the activation of EGF/ErbB receptors. Mol Cell 2003;12:541-552.

30. Stamos J, Sliwkowski MX, Eigenbrot C. Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. J Biol Chem. 2002;277:46265-46272.

31. Lorenzato A, Olivero M, Patrane S, et al. Novel somatic mutations of the MET oncogene in human carcinoma metastases activating cell motility and invasion. Cancer Res 2002; 62:7025-30.

32. Davies H, Bignell GR, Cox C, et al. Mutations of the BRAF gene in human cancer. Nature 2002;417:906-7.

33. Bardelli A, Parsons DW, Silliman N, et al. Mutational analysis of the tyrosine kinome in colorectal cancers. Science 2003;300:949.

34. Daley GQ, Van Etten RA, Baltimore D. Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome. Science 1990;247:824-30.

35. Heinrich, MC, Corless CL, Demetri GD, et al. Kinase mutations and imatinib response in patients with metastatic gastrointestinal stromal tumor. J Clin Oncol 2003;21:4342-4349.

36. Li B, Chang C, Yuan M, McKenna WG, Shu HG. Resistance to small molecule inhibitors of epidermal growth factor receptor in malignant gliomas. Cancer Res 2003;63:7443-7450.

37. C. L. Sawyers, Genes Dev 17, 2998-3010 (2003).

38. G. D. Demetri et al., N Engl J Med 347, 472-80 (2002).

39. B. J. Druker et al., N Engl J Med 344, 1038-42. (2001).

40. D. J. Slamon et al., N Engl J Med 344, 783-92 (2001).

41. H. Davies et al., Nature 417, 949-54 (2002).

42. Bardelli et al., Science 300, 949 (2003).

43. Y. Samuels et al., Science (2004).

44. Jemal et al., CA Cancer J Clin 54, 8-29 (2004).

45. S. Breathnach et al., J Clin Oncol 19, 1734-1742 (2001).

46. V. Rusch et al., Cancer Res 53, 2379-85 (1993).

47. R. Bailey et al., Lung Cancer 41 S2, 71 (2003).

48. M. Fukuoka et al., J Clin Oncol 21, 2237-46 (2003).

49. P. A. Janne et al., Lung Cancer 44, 221-230 (2004).

50. M. G. Kris et al., Jama 290, 2149-58 (2003).

51. V. A. Miller et al., J Clin Oncol 22, 1103-9 (2004).

52. M. Huse, J. Kuriyan, Cell 109, 275-82 (2002).

53. K. Naoki, T. H. Chen, W. G. Richards, D. J. Sugarbaker, M. Meyerson, Cancer Res 62, 7001-3 (2002).

54. J. Stamos, M. X. Sliwkowski, C. Eigenbrot, J Biol Chem 277, 46265-72 (2002).

55. T. Fujishita et al., Oncology 64, 399-406 (2003).

56. M. Ono et al., Mol Cancer Ther 3, 465-472 (2004).

57. M. C. Heinrich et al., J Clin Oncol 21, 4342-9 (2003).

58. G. Giaccone et al., J Clin Oncol 22, 777-84 (2004).

59. R. S. Herbst et al., J Clin Oncol 22, 785-94 (2004).

60. H. Yamazaki et al., Mol Cell Biol 8, 1816-20 (1988).

61. M. E. Gorre et al., Science 293, 876-80 (2001).

62. Marchetti A, Martella C, Felicioni L, et al: EGFR mutations in non-small-cell lung cancer: analysis of a large series of cases and development of a rapid and sensitive method for diagnostic screening with potential implications on pharmacologic treatment. J Clin Oncol 23:857-65, 2005.

63. Franklin WA: Diagnosis of lung cancer: pathology of invasive and preinvasive neoplasia. Chest 117:80S-89S, 2000.

64. Paez JG, Janne PA, Lee JC, et al: EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science 304:1497-500, 2004.

65. Lynch TJ, Bell DW, Sordella R, et al: Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. N Engl J Med 350:2129-39, 2004.

66. Pao W, Miller V, Zakowski M, et al: EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. Proc Natl Acad Sci U S A 101:13306-11, 2004

67. Huang SF, Liu HP, Li LH, et al: High frequency of epidermal growth factor receptor mutations with complex patterns in non-small cell lung cancers related to gefitinib responsiveness in Taiwan. Clin Cancer Res 10:8195-203, 2004.

68. Han SW, Kim TY, Hwang PG, et al: Predictive and Prognostic Impact of Epidermal Growth Factor Receptor Mutation in Non-Small-Cell Lung Cancer Patients Treated With Gefitinib. J Clin Oncol, 2005.

69. Tokumo M, Toyooka S, Kiura K, et al: The relationship between epidermal growth factor receptor mutations and clinicopathologic features in non-small cell lung cancers. Clin Cancer Res 11:1167-73, 2005.

70. Mitsudomi T, Kosaka T, Endoh H, et al: Mutations of the Epidermal Growth Factor Receptor Gene Predict Prolonged Survival After Gefitinib Treatment in Patients with Non-Small-Cell Lung Cancer With Postoperative Recurrence. J Clin Oncol, 2005.

71. Pao W, Wang TY, Riely GJ, et al: KRAS Mutations and Primary Resistance of Lung Adenocarcinomas to Gefitinib or Erlotinib. PLoS Med 2:e17, 2005

72. Read WL, Page NC, Tierney RM, et al: The epidemiology of bronchioloalveolar carcinoma over the past two decades: analysis of the SEER database. Lung Cancer 45:137-42, 2004.

73. Sanderson Cox L, Sloan JA, Patten CA, et al: Smoking behavior of 226 patients with diagnosis of stage IIIA/IIIB non-small cell lung cancer. Psychooncology 11:472-8, 2002.

74. Radzikowska E, Glaz P, Roszkowski K: Lung cancer in women: age, smoking, histology, performance status, stage, initial treatment and survival. Population-based study of 20 561 cases. Ann Oncol 13:1087-93, 2002.

75. Tong L, Spitz MR, Fueger JJ, et al: Lung carcinoma in former smokers. Cancer 78:1004-10, 1996.

76. de Perrot M, Licker M, Bouchardy C, et al: Sex differences in presentation, management, and prognosis of patients with non-small cell lung carcinoma. J Thorac Cardiovasc Surg 119:21-6, 2000

77. Capewell S, Sankaran R, Lamb D, et al: Lung cancer in lifelong non-smokers. Edinburgh Lung Cancer Group. Thorax 46:565-8, 1991

78. Gritz ER, Nisenbaum R, Elashoff RE, et al: Smoking behavior following diagnosis in patients with stage I non-small cell lung cancer. Cancer Causes Control 2:105-12, 1991

79. Sridhar KS, Raub WA, Jr.: Present and past smoking history and other predisposing factors in 100 lung cancer patients. Chest 101:19-25, 1992

80. Barbone F, Bovenzi M, Cavallieri F, et al: Cigarette smoking and histologic type of lung cancer in men. Chest 112:1474-9, 1997

81. Shigematsu H, Lin L, Takahashi T, et al: Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. J Natl Cancer Inst 97:339-46, 2005

82. Kosaka T, Yatabe Y, Endoh H, et al: Mutations of the epidermal growth factor receptor gene in lung cancer: biological and clinical implications. Cancer Res 64:8919-23, 2004

83. Cho D, Kocher O, Tenen DG, et al: Unusual cases in multiple myeloma and a dramatic response in metastatic lung cancer: case 4. Mutation of the epidermal growth factor receptor in an elderly man with advanced, gefitinib-responsive, non-small-cell lung cancer. J Clin Oncol 23:235-7, 2005

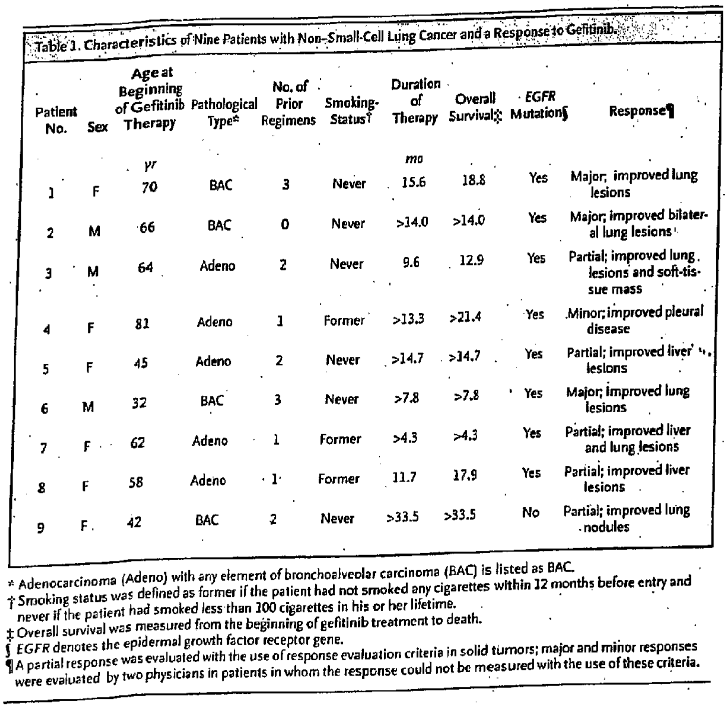


Table 1. Characteristics of Nine Patients with Non–Small-Cell Lung Cancer and a Response to Gefitinib.

Patient No.	Sex	Age at Beginning of Gefitinib Therapy	Pathological Type*	No. of Prior Regimens	Smoking Status†	Duration of Therapy	Overall Survival‡	*EGFR* Mutation§	Response¶
		yr				*mo*			
1	F	70	BAC	3	Never	15.6	18.8	Yes	Major; improved lung lesions
2	M	66	BAC	0	Never	>14.0	>14.0	Yes	Major; improved bilateral lung lesions
3	M	64	Adeno	2	Never	9.6	12.9	Yes	Partial; improved lung lesions and soft-tissue mass
4	F	81	Adeno	1	Former	>13.3	>21.4	Yes	Minor; improved pleural disease
5	F	45	Adeno	2	Never	>14.7	>14.7	Yes	Partial; improved liver lesions
6	M	32	BAC	3	Never	>7.8	>7.8	Yes	Major; improved lung lesions
7	F	62	Adeno	1	Former	>4.3	>4.3	Yes	Partial; improved liver and lung lesions
8	F	58	Adeno	1	Former	11.7	17.9	Yes	Partial; improved liver lesions
9	F	42	BAC	2	Never	>33.5	>33.5	No	Partial; improved lung nodules

* Adenocarcinoma (Adeno) with any element of bronchoalveolar carcinoma (BAC) is listed as BAC.

† Smoking status was defined as former if the patient had not smoked any cigarettes within 12 months before entry and never if the patient had smoked less than 100 cigarettes in his or her lifetime.

‡ Overall survival was measured from the beginning of gefitinib treatment to death.

§ *EGFR* denotes the epidermal growth factor receptor gene.

¶ A partial response was evaluated with the use of response evaluation criteria in solid tumors; major and minor responses were evaluated by two physicians in patients in whom the response could not be measured with the use of these criteria.

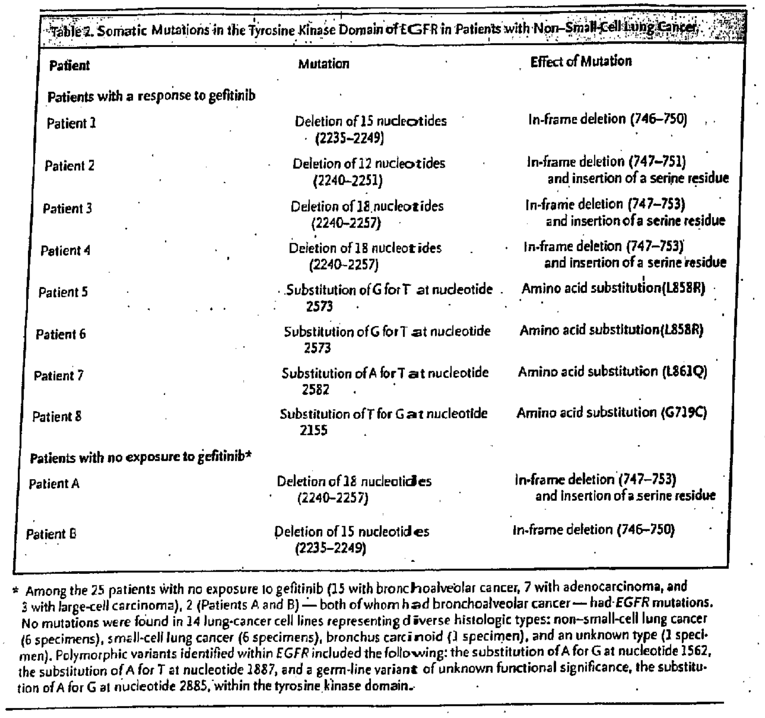


Table 2. Somatic Mutations in the Tyrosine Kinase Domain of *EGFR* in Patients with Non–Small-Cell Lung Cancer.

Patient	Mutation	Effect of Mutation
Patients with a response to gefitinib		
Patient 1	Deletion of 15 nucleotides (2235–2249)	In-frame deletion (746–750)
Patient 2	Deletion of 12 nucleotides (2240–2251)	In-frame deletion (747–751) and insertion of a serine residue
Patient 3	Deletion of 18 nucleotides (2240–2257)	In-frame deletion (747–753) and insertion of a serine residue
Patient 4	Deletion of 18 nucleotides (2240–2257)	In-frame deletion (747–753) and insertion of a serine residue
Patient 5	Substitution of G for T at nucleotide 2573	Amino acid substitution (L858R)
Patient 6	Substitution of G for T at nucleotide 2573	Amino acid substitution (L858R)
Patient 7	Substitution of A for T at nucleotide 2582	Amino acid substitution (L861Q)
Patient 8	Substitution of T for G at nucleotide 2155	Amino acid substitution (G719C)
Patients with no exposure to gefitinib*		
Patient A	Deletion of 18 nucleotides (2240–2257)	In-frame deletion (747–753) and insertion of a serine residue
Patient B	Deletion of 15 nucleotides (2235–2249)	In-frame deletion (746–750)

* Among the 25 patients with no exposure to gefitinib (15 with bronchoalveolar cancer, 7 with adenocarcinoma, and 3 with large-cell carcinoma), 2 (Patients A and B) — both of whom had bronchoalveolar cancer — had *EGFR* mutations. No mutations were found in 14 lung-cancer cell lines representing diverse histologic types: non–small-cell lung cancer (6 specimens), small-cell lung cancer (6 specimens), bronchus carcinoid (1 specimen), and an unknown type (1 specimen). Polymorphic variants identified within *EGFR* included the following: the substitution of A for G at nucleotide 1562, the substitution of A for T at nucleotide 1887, and a germ-line variant of unknown functional significance, the substitution of A for G at nucleotide 2885, within the tyrosine kinase domain.

127

Table 4: Population Characteristics Among 100 Patients Tested for *EGFR* Mutations as Part of NSCLC Care

Characteristic	Frequency
Mean age, years (standard deviation)	60.7 (11.0)
Female	63
Race	
White	76
Asian	7
Other	5
Unknown	12
Stage at Time of Test	
I	15
II	4
III	10
IV	67
Unknown	4
Histology	
Pure BAC	1
Adenocarcinoma with BAC Features	24
Adenocarcinoma	69
NSCLC, all other subtypes	6
Smoking Status	
Current	17
Former	48
Never	29
Unknown	6
Mean amount smoked by current and former smokers, pack-years (standard deviation)	39.0 (32.3)
Mean time from diagnosis to *EGFR* test, months (standard deviation)	18.7 (78.4)
Prior Chemotherapy Treatment	47
Prior EGFR Targeted Treatment	11

BAC = Bronchioloalveolar Carcinoma, EGFR = Epidermal Growth Factor Receptor

Table 5: Epidermal Growth Factor Receptor Somatic Gene Mutations Identified

Patient	Gender	Histology	Pack-Years Smoked	Exon	Nucleotide Change	Amino Acid Change
1	F	Adeno	0	18 18	2126A>T 2155G>A	E709V G719S
2	F	A+BAC	60	18 20	2156G>C 2327G>A	G719A R776H
3	F	A+BAC	0	19	2235_2249 del	K745_A750 del ins K
4	M	A+BAC	0	19	2235_2249 del	K745_A750 del ins K
5	F	Adeno	5	19	2235_2249 del	K745_A750 del ins K
6	M	Adeno	Unknown	19	2235_2249 del	K745_A750 del ins K
7	F	Adeno	0	19	2236_2250 del	E746_A750 del
8	M	Adeno	45	19	2236_2250 del	E746_A750 del
9	F	Adeno	Unknown	19	2236_2250 del	E746_A750 del
10	M	A+BAC	12	19	2237_2255 del ins T	E746_S752 del ins V
11	M	Adeno	1	19	2239_2248 del ins C	L747_A750 del ins P
12	M	A+BAC	0	19	2239_2251 del ins C	L747_T751 del ins P
13	F	Adeno	30	19	2253_2276 del	T751_I759 del ins T
14	F	Adeno	0	19	2254_2277 del	S752_I759 del
15	F	Adeno	0	20	2303_2311 dup	D770_N771 ins SVD
16	M	Adeno	5	20	2313_2318 dup CCCCCA	P772_H773 dup
17	F	Adeno	0	21	2543C>T	P848L*
18	M	BAC	0	21	2573T>G	L858R
19	F	A+BAC	0	21	2573T>G	L858R
20	M	A+BAC	1	21	2573T>G	L858R
21	F	Adeno	0	21	2573T>G	L858R
22	F	Adeno	15	21	2573T>G	L858R
23	F	Adeno	0	21	2582T>A	L861Q

Adeno = Adenocarcinoma, Adeno + BAC = Adenocarcinoma with Bronchioloalveolar Carcinoma Features, BAC = Pure Bronchioloalveolar Carcinoma

* This mutation was identified as a germline variant

Table S1A: Primers for amplification of selected EGFR and receptor tyrosine kinase exons (SEQ ID NOS: 1-212)

Gene	RefSeq	Exon	SEQ ID NO	F Nested	R Nested
ALK	NM_004304	24	1,2	GGAAATATAGGGAAGGGAAGGAA	TTGACAGGGTACCAGGGAGATGA
ALK	NM_004304	25	3,4	CTGAACCGCCAAGGACTCAT	TTTTCCCTCCCTACTAACACACG
AXL	NM_021913	19	5,6	ACTGATGCCCTGACCCTGTT	CCCATGGTTCCCCACTCCTT
CSF1R	NM_005211	18	7,8	AGGGACTCCAAAGCCATGTG	CTCTCTGGGGCCATCCAACT
CSF1R	NM_005211	19	9,10	CATTGTCAAGGGCAATGTAAGTG	CTCTCACCAACCCTCGCCTGT
DDR1	NM_013994	15	11,12	ACATGGGGAGCCAGAGTGAC	TGCAACCCAGAGAAAGTGTG
DDR2	NM_006182	16	13,14	TGAGCTTTCAACCCTAGTTTGTTG	GTTTGCCTCCTGCTGTCCTCA
DKFZp761P1010	NM_018423	8	15,16	TGTCCTTGTGTTTTTGAAGATTCC	TGCAGACAGATGACAAACATGAA
EGFR	NM_005228	2	17,18	TGGGTGAGTCTCTGTGTGGAG	CATTGCCATAGCAAAAATAAACACA
EGFR	NM_005228	3	19,20	GGTTCAACTGGGCGTCCTA	CCTTCTCCGAGGTGGAATTG
EGFR	NM_005228	4	21,22	CGCACCATGGCATCTCTTTA	AAAACGATCTCTATGTCCCGTGGT
EGFR	NM_005228	5	23,24	CAGCCAGCCAAACAATCAGA	TCTTTGGAGTCTTCAGAGGGAAA
EGFR	NM_005228	6	25,26	TGTGGTTTCGTTGGAAGCAA	AATTGACAGCTCCCCCAACAG
EGFR	NM_005228	7	27,28	GGCTTTCTGACGGGAGTCAA	CCACCCAAAGACTCTCCCAAGA
EGFR	NM_005228	8	29,30	CCTTTCCATCACCCCTCAAG	AGTGCCTTCCCATTGCCCTAA
EGFR	NM_005228	9	31,32	ACCGGAATTCCTTCCTGCTT	CACTGAAACAAACAACCAGGGTGA
EGFR	NM_005228	10	33,34	AGGGGGTGAGTCACAGGTTC	TCAGAAGAAATGTTTTTTATTCCAAGG
EGFR	NM_005228	11	35,36	GCAAATCCAATTTTCCCACTT	GCAGGAGCTCTGTGCCCCTAT
EGFR	NM_005228	12	37,38	TCCCACAGCATGACCTACCA	TTTGCTTCTTAAGGAACCTGAAAA
EGFR	NM_005228	13	39,40	TGTCACCCAAGGTCATGGAG	CAAAAGCCAAGGGCAAAGAA
EGFR	NM_005228	14	41,42	GGAGTCCCAACTCCTTGACC	GTCCTGCCCACACAGGAATG
EGFR	NM_005228	15	43,44	GCTTTCCCCACTCACACACA	CAAACCTCGGCAATTTGGTTG
EGFR	NM_005228	16	45,46	CCACCAATCCAACATCCAGA	TGGCCCAGAGCCATAGAAAC
EGFR	NM_005228	17	47,48	TTCCAAGATCATTCTACAAGATGTCA	GCACATTCAGAGATTCTTTTCTGC
EGFR	NM_005228	18	49,50	TCCAAATGAGCTGGCAAGTG	TCCCAAACACTCAGTGAAAACAAA
EGFR	NM_005228	19	51,52	GTGCATCGCTGGTAACATCC	TGTGGAGATGAGCAGGGTCT
EGFR	NM_005228	20	53,54	ATCGCATTCATGCGTCTTCA	ATCCCCATGGCAAACTCCTTG
EGFR	NM_005228	21	55,56	GCTCAGAGCCTGGCATGAA	CATCCTCCCCTGCATGTTGT
EGFR	NM_005228	22	57,58	TGGCTCGTCTGTGTGTGTCA	CGAAAGAAAATACTTGCCATGTCAGA
EGFR	NM_005228	23	59,60	TGAAGCAAATTGCCCAAGAC	TGACATTTCTCCAGGGATGC
EGFR	NM_005228	24	61,62	AAGTGTCGCATCACCAATGC	ATGCGATCTGGGACACAAGG
EGFR	NM_005228	25	63,64	GGCACCTGCTGGCAATAGAC	TGACTTCATATCCATGTTGAGTTTCACT
EGFR	NM_005228	26	65,66	TATACCCTCCATGAGGCACA	GGGAAAAACCCACACAAGGAA
EGFR	NM_005228	27	67,68	TCAGAACCAGCATCTCAAGGA	GATGCTGGAGGGAGCACCCT
EGFR	NM_005228	28 1	69,70	CCTTGTTGAGGACATTCACAGG	ATGTGCCCGAGGTGGAAGTA
EPHA1	NM_005232	14	71,72	GGAGGGCAGAGGACTAGCTG	GTGCCTGGCCAAGTCTTTGT
EPHA1	NM_005232	15	73,74	CTGCAGCCTAGCAACAGAGC	AAGAACCAGAGGAGCCCAGGA
EPHA2	NM_004431	13	75,76	CGGGTAAGGATGTGGGTTGT	CAGGTGTTCTGCCTCCTTGAA
EPHA2	NM_004431	14	77,78	GCTTCAGGAGGCAGAACACC	GGAGCAAGCCTAAGAACGGTTCA
EPHA3	NM_005233	10	79,80	GCCTTGTATCCATTTGCCACA	TGACAACACGTTTTGGGTCAT
EPHA3	NM_005233	11	81,82	TGCATATTCCATTTCAGAACAGA	AAACAGTTTCATTGCTGCTAAAT
EPHA4	NM_004438	13	83,84	CCGGATACAGATACCCAAAAAGA	GGAGGCTTCAAGGGATGAGA
EPHA4	NM_004438	14	85,86	GCTGTTGTCCTGCTTGGCTA	TGGTTGTAATGTTGAACTAGCTTGC
EPHA7	NM_004440	13	87,88	TGGCTGTCAGCTAAATAAGCATGT	TCAATTTGCTTCATTTCTTCCTGTT
EPHA7	NM_004440	14	89,90	TGCTGCTGAACTACCAACCAA	TGTGGTAGTAATTGTGGAAAACTG
EPHA8	NM_020526	13	91,92	CAAAGCACCGTCTCAACTCG	CCCGAAACTGCCAACTTCAT
EPHA8	NM_020526	14	93,94	GGAAAACAGGACCCCAGTGT	CCCTCCTCCACAGAGCTTGAT
EPHB1	NM_004441	7	95,96	GACAGAAGCTGACAAGCAGCA	AGGTTCCATTCCCTCCCAAGT
EPHB1	NM_004441	8	97,98	TGGGAGTGAGAGTTTGGAAGAA	TATGAGGCCGTGAGCTGAAA
EPHB2	NM_017449	11	99,100	AGGGCCCTGCTCTGGTTT	CCAATTGGGCGTTAGTGAAA
EPHB2	NM_017449	12	101,	CTCATGAGATTGGGGCATCA	AGGCCCATGATCTCAGAAGC
EPHB3	NM_004443	11	103,	GGTTGCAGGAGAGACGAGGT	AGGCCCTTCACCCTGTGAAC
EPHB3	NM_004443	12	105	ATGACCCCTCCGATCCTACC	TAATCCTGCTCCACGGCAATT

Table S1A: Continued

EPHB4	NM_004444	14	107	GGAAAAAGCAGAGGCAGGTG	TGGTCTCAAGAACCCAGCCAG
EPHB6	NM_004445	16	109	GACACCCTCCCCCTCTCAT	ACTATGACACCCCGGCTCGAG
EPHB6	NM_004445	17	111	TGCTTGATGTAAAACCCTTGG	GCAATCCAACAGCCATGAGA
ERBB2	NM_004448	21	113	GGAGCAAACCCCTATGTCCA	TCCTCCAACTGTGTGTTCGTGG
ERBB3	NM_001982	21	115	TGGGGACCACTGCTGAGAG	TGCAGCCTTCTCTCCTTGAA
FGFR1	NM_000604	14	117	GCAGAGCAGTGTGGCAGAAG	ACAGGTGGGAAGGGACTGG
FGFR1	NM_000604	15	119	AGTGGGGTGGGCTGAGAAC	TCTCTGGGGCAGAAAGAGGA
FGFR2	NM_000141	14	121	ACCCGGCCACACTGTATTTC	CATCCCACCCAGCTCTCAAC
FGFR2	NM_000141	15	123	AGGGCATAGCCCTATTGAGC	CCCAGGAAAAAGCCAGAGAA
FGFR3	NM_000142	13	125	CAGGTGTGGGTGGAGTAGGC	CTCAGGCGCCATCCACTT
FGFR3	NM_000142	14	127	AAGAAGACGACCAACGTGAGC	AGGAGCTCCAGGGCACAG
FGFR4	NM_002011	14	129	CCTCCTCTGTAAAGTGGGTGGA	AGAGGGCCTCAGTGCAGAGT
FGFR4	NM_002011	15	131	AGATGGGGCAGAACTGGATG	GGGTCCCAGACCAAATCCTGA
FLT1	NM_002019	23	133	AGGTGCTCCCTTCACAGCAT	TTCAGGGACTACAGCTGAGGAA
FLT1	NM_002019	24	135	GCCGTATGTTATCTGGGAGGT	TGGGCCCATTACACTTTAAGA
FLT3	NM_004119	20	137	TTCCATCACCGGTACCTCCT	CCATAAATCAAAAATGCACCACA
FLT3	NM_004119	21	139	GAGTGGTCTTAGGAAGATGATGC	AAAGTCATGGGCTGCAATACAA
FLT4	NM_002020	23	141	ATGGTCCCCACTGCTTGG	AGGAGCTCACCTCACCCTGT
IGF1R	NM_000875	18	143	CCTTGCGTCTCTCCACACAT	TGGCAACGGGTAACAATGAA
INSR	NM_000208	18	145	GGCTGAGGTAAGCTGCTTCG	AAAAAGAAGTATCTTGCCCCCTTT
INSR	NM_000208	19	147	AACCCCTCTTAGGGCTCTGTG	CAGGAGGATGGCAGGCTTTC
KDR	NM_002253	24	149	CGTAGAGAGCTTCAGGACCTGTG	TTCCGAGAAGTTTTGCCTGA
KIT	NM_000222	17	151	TGTGAACATCATTCAAGGCGTA	AAAATGTGTGATATCCCTTAGACAGG
KIT	NM_000222	18	153	TCCACATTTCAGCAACAGCA	GGCTGCTTCCTGAGACACCAGT
LTK	NM_002344	16	155	TATCTACCGGTGCGGGACTT	AGGTGTAGCCTCCCCTCACA
MERTK	NM_006343	17	157	AGGCTGGTGGTGTCTCTGTG	CAAGCTGCCAACCCTCACGTT
MET	NM_000245	19	159	TGGATTTCAAATACTGAAGCCACT	TGGAATTGGTGGTGTTGAATTT
MUSK	NM_005592	15_1	161	GGGCTTCATATGTTCTGACATGG	CAGAGGACCACGCCATAGG
MUSK	NM_005592	15_2	163	CCGAGATTTAGCCACCAGGA	CCTGGGAAGCAAACAACCACA
NTRK1	NM_002529	15	165	AGGTCCCCAGTCTCCTCTCC	AGACCCATGCAGCCATCCTA
NTRK1	NM_002529	16	167	CGTGAACCACCGAGCTTGT	AGAGGGGCAGAAGGGGAAC
NTRK2	NM_006180	15	169	GGTGGGGGTGAGGAGCTTAG	TCGTTTAAGCCACCCAGTTCA
NTRK2	NM_006180	16	171	TGCAAATAAGGAAAGCAAACA	TCCTGACATGGTCTTCCAACC
NTRK3	NM_002530	17	173	CAGCATCTTCACACACCTCTGA	GCTGGCTCTAAATCCCACCT
NTRK3	NM_002530	18	175	CTAATCCGGGAAGTTGTTGC	TTCTGTATCAGCAGCTTCCTCTGTG
PDGFRA	NM_006206	18	177	CAAGTGCCACCATGGATCA	GGCAGTGTACTGACCCCTTGA
PDGFRA	NM_006206	19	179	GCACAAGTTATTAAGAGCCCAAGG	AGCATACTGGCCTCACACCA
PDGFRB	NM_002609	18	181	GCACATGGGCAGTGTTGTATTT	GAGCCCCACACAGATTTCCT
PDGFRB	NM_002609	19	183	ATGGGACGGAGAAGTGGTTG	TCCCTGTATCAGGGCTCCGTC
PTK7	NM_002821	18	185	TTCCTACGCAGCACACCAAT	GCAGGCACTAAACCCTTTCC
PTK7	NM_002821	19	187	GCACGCATGTGACCAATTTC	AGCCCTGAGAGGGAGGTAGG
RET	NM_000323	15	189	CACACACCACCCCTCTGCT	AAAGATTTGGGGTGAGGGCTA
RET	NM_000323	16	191	CTGAAAGCTCAGGGATAGGG	CTGGCCAAGCTGCACAGA
ROR1	NM_005012	09_1	193	TGCAGCCAACGATTTGAAAG	GGAAAGCCCCAAGTCTGAAA
ROR1	NM_005012	09_2	195	TCATCATGAGATCCCCACACT	GCATTTCCCCCTGAAGGAGT
ROR1	NM_005012	09_3	197	TGGATTCAGTAACCAGGAAGTGA	CCCATTCCACCAGGATGATT
ROR1	NM_005012	09_4	199	GTTTCCAGCTGCCCACTACC	GCTCGAAACCACATGTTCCA
RYK	NM_002958	13	201	CTGGATTTGGGGTTCTCTGC	CGGGAACAGCTAGCAGATTTTT
TEK	NM_000459	18	203	GGGAATTTTGGAGGGGAACT	GCTTCAGTCACCACAGACGCA
TEK	NM_000459	19	205	TGAGTCTACCCAGCAATCATTTG	TTCCCGAGAGCTACAGGACA
TIE	NM_005424	18	207	GGTAACAAGGGTACCCACGAA	GTTTGAGGGGCTGAGTGTGG
TIE	NM_005424	19	209	CCTCACCCTTAGGGCTTGTG	AGCCCAGGTCATGCCTTAGA
TYRO3	NM_006293	18	211,212	GGGTAGCTTGGGAGCAAAGA	CCAAACCCCAGAGAGCAGAC

Table S1B: Primers for amplification of selected EGFR and receptor tyrosine kinase exons (SEQ ID NOS: 213 - 424)

Gene	RefSeq	Exon	SEQ ID NO	F External	R External
ALK	NM_004304	24	213,214	CATTTCCCCTAATCCTTTTCCA	GTGATCCCAGATTTAGGCCTTC
ALK	NM_004304	25	215	GCCTCTCGTGGTTTGTTTTGTC	CCCAGGGTAGGGTCCAATAATC
AXL	NM_021913	19	217	CTTCCTGGTGGAGGTGACTGAT	CAGGCATAGTGTGTGATGGTCA
CSF1R	NM_005211	18	219	TCACGATACACATTCTCAGATCC	GAAGATCTCCCAGAGGAGGATG
CSF1R	NM_005211	19	221	CGTAACGTGCTGTTGACCAAT	AAACGAGGGAAGAGCCAGAAAG
DDR1	NM_013994	15	223	TGGGGAGCACAATAAAAGAAGA	ACTCTTGGCTCCTGGATTCTTG
DDR2	NM_006182	16	225	GGAAGTCAGTGTGCAGGGAATA	TTTTAGCAGAAATAGGCAAGCA
DKFZp761P1010	NM_018423	8	227	TGGTAATCCTAAACACAATGCAGA	CTGGGCAACACAGTGAGATCCT
EGFR	NM_005228	2	229	TCACAAATTTCTTTGCTGTGTCC	CATGGAACTCCAGATTAGCCTGT
EGFR	NM_005228	3	231	GATTGTTGCAGATCGTGGACAT	CGCTTAAATCTTCCCATTCCAG
EGFR	NM_005228	4	233	CTCCATGGCACCATCATTAACA	CTCAGGACACAAGTGCTCTGCT
EGFR	NM_005228	5	235	GCAGTTCATGGTTCATCTTCTTTT	CAAAATAGCCCACCCTGGATTA
EGFR	NM_005228	6	237	CTTTCTGCATTGCCCAAGATG	CAAGGTCTCAGTGAGTGGTGGA
EGFR	NM_005228	7	239	GAGAAGGGTCTTTCTGACTCTGC	CAGGTGTTTCTCCTGTGAGGTG
EGFR	NM_005228	8	241	CACATTGCGGCCTAGAATGTTA	ACCCCGTCACAACCTTCAGT
EGFR	NM_005228	9	243	GCCGTAGCCCCAAAGTGTACTA	TCAGCTCAAACCTGTGATTTCC
EGFR	NM_005228	10	245	CTCACTCTCCATAAATGCTACGAA	GACTTAACGTGTCCCCTTTTGC
EGFR	NM_005228	11	247	GCCTCTTCGGGGTAATCAGATA	GAAGTCTGTGGTTTAGCGGACA
EGFR	NM_005228	12	249	ATCTTTTGCCTGGAGGAACTTT	CAGGGTAAATTCATCCCATTGA
EGFR	NM_005228	13	251	CAGCAGCCAGCACAACTACTTT	TTGGCTAGATGAACCATTGATGA
EGFR	NM_005228	14	253	TGAATGAAGCTCCTGTGTTTACTC	ATGTTCATCGCAGGCTAATGTG
EGFR	NM_005228	15	255	AAAACAGGGAGAACTTCTAAGCAA	CATGGCAGAGTCATTCCCACT
EGFR	NM_005228	16	257	CAATGCTAGAACAACGCCTGTC	TCCCTCCACTGAGGACAAAGTT
EGFR	NM_005228	17	259	GGGAGAGCTTGAGAAAGTTGGA	ATTTCCTCGGATGGATGTACCA
EGFR	NM_005228	18	261	TCAGAGCCTGTGTTTCTACCAA	TGGTCTCACAGGACCACTGATT
EGFR	NM_005228	19	263	AAATAATCAGTGTGATTCGTGGAG	GAGGCCAGTGCTGTCTCTAAGG
EGFR	NM_005228	20	265	ACTTCACAGCCCTGCGTAAAC	ATGGGACAGGCACTGATTTGT
EGFR	NM_005228	21	267	GCAGCGGGTTACATCTTCTTTC	CAGCTCTGGCTCACACTACCAG
EGFR	NM_005228	22	269	CCTGAACTCCGTCAGACTGAAA	GCAGCTGGACTCGATTTCCT
EGFR	NM_005228	23	271	CCTTACAGCAATCCTGTGAAACA	TGCCCAATGAGTCAAGAAGTGT
EGFR	NM_005228	24	273	ATGTACAGTGCTGGCATGGTCT	CACTCACGGATGCTGCTTAGTT
EGFR	NM_005228	25	275	TAAGGCACCCACATCATGTCA	TGGACCTAAAAGGCTTACAATCAA
EGFR	NM_005228	26	277	GCCTTTTAGGTCCACTATGGAATG	CCAGGCGATGCTACTACTGGTC
EGFR	NM_005228	27	279	TCATAGCACACCTCCCTCACTG	ACACAACAAAGAGCTTGTGCAG
EGFR	NM_005228	28 1	281	CCATTACTTTGAGAAGGACAGGAA	TATTCTTGCTGGATGCGTTTCT
EPHA1	NM_005232	14	283	AGGAGGGCAGAGGACTAGCTG	GGCAATGTGAATGTGCACTG
EPHA1	NM_005232	15	285	CTTGAACCTGGGAGGTGGAG	ATCAGGGTGGGAGGAGTAAAGA
EPHA2	NM_004431	13	287	CCCACTTACCTCTCACCTGTGC	GTGAACTTCCGGTAGGAAATGG
EPHA2	NM_004431	14	289	AGGGGACCTCAAGGGAGAAG	AGATCATGCCAGTGAACTCCAG
EPHA3	NM_005233	10	291	GGACCAGGAAAGTCCTTGCTTT	GGTGGGGAACATTAAACTGAGG
EPHA3	NM_005233	11	293	GCTTCAGGTTGTTTTGTTGCAG	ACCCTTGCTTGAGGGAAATATG
EPHA4	NM_004438	13	295	CCCAGCTCCTAGGGTACAGTCT	CAGTCAGCTTCAAAATCCCTCTT
EPHA4	NM_004438	14	297	TCACTTCCCTGTGAGTAAAGAAAA	GGCCATTTAATTCTTGTCCTTGA
EPHA7	NM_004440	13	299	TGGACTTGTGCAAACTCAAACTG	TCCCAATATAGGGCAGTCATGTT
EPHA7	NM_004440	14	301	TCTCAATCAGTTGAGTTGCCTTG	AGCTGTGCAAGTGTGGAAACAT
EPHA8	NM_020526	13	303	GCTGTGAGGGTAAATGAGACCA	GTCTCCTGGTGAGTGACTGTGG
EPHA8	NM_020526	14	305	CCTTCCTTCGTCTCCACAGC	GTCCTTGTGCCAACAGTCGAG
EPHB1	NM_004441	7	307	GCTTGGCAAGGAGAAGAGAACA	GCTTGCTTTCTTGCTTGAACAAC
EPHB1	NM_004441	8	309	GCTGGTCACCTTGAGCTTCTCT	CCATGCTGGGCTCTTTGATTA
EPHB2	NM_017449	11	311	CACCACTCTGAAGTTGGCCTCT	ATGGCTCTGCACATTTGTTCC
EPHB2	NM_017449	12	313	CAGAGTGGGAAAAGGCACTTCA	CCAGAGTCCTGTGCAGACATTC
EPHB3	NM_004443	11	315	ATGGGGATTAACTGGGATGTTG	CGTAGCTCCAGACATCACTAGCA
EPHB3	NM_004443	12	317	GCAACCTGGTCTGCAAAGTCTC	ACCCAGCAGTCCAGCATGAG

TABLE SIB CONT.

EPHB4	NM 004444	14	319	GAGTTTCAGTGAGCCAAGATCG	TTACAGGCTTGAGCCACTAGGC
EPHB6	NM 004445	16	321	AAGCTTCCAGGAGACGAGGTC	GTCCCTGAAATCCCTCAAACC
EPHB6	NM 004445	17	323	TGCTCCATAAACGTGACTATTGC	GTAAGAGGGTGGGCTGGAATCT
ERBB2	NM 004448	21	325	CTTAGACCATGTCCGGGAAAAC	CACATCACTCTGGTGGGTGAAC
ERBB3	NM 001982	21	327	AAATTTCATCCCAAAACCAACC	CCAGTCCCAAGTTCTTGATCATT
FGFR1	NM 000604	14	329	ACAAGTCGGCTAGTTGCATGG	TCTCAGATGAAACCACCAGCAC
FGFR1	NM 000604	15	331	TTCATCTGAGAAGCAAGGAGTGG	CCAGGGAGAAAGCAGGACTCTA
FGFR2	NM 000141	14	333	TTCTGGCGGTGTTTTGAAATTA	CTCAACATTGACGGCCTTTCTT
FGFR2	NM 000141	15	335	TCAGCTCTTAAACAGGGCATAGC	GAAATGCAGCAGCCACTAAAGA
FGFR3	NM 000142	13	337	CTCACCTTCAAGGACCTGGTGT	CAGGGAGGGGTAGAAACCACA
FGFR3	NM 000142	14	339	GGAGAGGTGGAGAGGCTTCAG	GAGACTCCCAGGACAGACACCT
FGFR4	NM 002011	14	341	CACTCGTTCCTCACCCTTCC	AGGACTCACACGTCACTCTGGT
FGFR4	NM 002011	15	343	GGACAATGTGATGAAGATTGCTG	ATAGCAGGATCCCAAAAGACCA
FLT1	NM 002019	23	345	GGCTTGGGGACCTGTATTTGTA	CAGTGGCCTTTCTGAGCCTTAC
FLT1	NM 002019	24	347	GCACTCTAGCTCCCTCTTTTAGC	TTTTACAGTAGAGGGCAGACATGC
FLT3	NM 004119	20	349	GCCACCATAGCTGCAGAATTAG	CCCAAGGACAGATGTGATGCTA
FLT3	NM 004119	21	351	GCCTTTGTTCGAGAGGAGTTGT	GTTCACGCTCTCAAGCAGGTTA
FLT4	NM 002020	23	353	ATTCCACAAGCTCTCTCCATGA	CTTGCCCCAAGATGCCTAAG
IGF1R	NM 000875	18	355	TGCTTGGTATTTGCTCATCATGT	CCCTTAGCTAGCCCACTGACAA
INSR	NM 000208	18	357	CTCCTGGGAGTGGTGTCCAA	CCTGGGCAACAGACAGAGTAAG
INSR	NM 000208	19	359	CTTCACTTCCCCATGCGTACC	GGGTTCACAATGCCTACAGGA
KDR	NM 002253	24	361	AAAATCTGTGACTTTGGCTTGG	GGGAGGAGACATTCTTTGATTTG
KIT	NM 000222	17	363	GCAGTCCTGAGAAGAAAACAGC	CTTCACATGCCCCAAAATTACA
KIT	NM 000222	18	365	TGAGCCATGTATTTCAGAGGTGA	TACATTTCAGCAGGTGCGTGTT
LTK	NM 002344	16	367	TTGCCTACTCTGTAGGGATATTGC	ATAGGGCATGTAGCCCAGTGA
MERTK	NM 006343	17	369	GCTCTGCTGTTGGTCCTCACT	TTGCAAAGCACACATCTTCTGA
MET	NM 000245	19	371	TGGCAATGTCAATGTCAAGCAT	GTATGTTGCCCCACTCAACAAA
MUSK	NM 005592	15 1	373	TGCATTTCCTAGCTGAGACTCC	TGCCATCTCGCACGTAGTAAAT
MUSK	NM 005592	15 2	375	CTCTCCTGTGCTGAGCAGCTTT	TGTTTCCAATCACTGGCTTTCA
NTRK1	NM 002529	15	377	GAACCATGGGCTGTCTCTGG	ATCTGGGATAGCGAAGGAGACA
NTRK1	NM 002529	16	379	ATTACAGGCCACACGCCATC	AAGGCAAGAATAAGGGAGGAAGA
NTRK2	NM 006180	15	381	GCTCTCAGGACTGCAGAAGTACA	GAGGAACCAATCCCACTCACAC
NTRK2	NM 006180	16	383	TCACTCTTTGCCTTCTGTCTCTG	GCACTGTGCTTTGCTTTCTCAG
NTRK3	NM 002530	17	385	TGTCTCCTTTATCGTAGGTCTCCA	CACCACATTTCCTACAGTTCCA
NTRK3	NM 002530	18	387	CACTGTGCACCAGACAGACAAA	TGTGGTTTTCTGTATCAGCAGCTT
PDGFRA	NM 006206	18	389	CAGGGAGTCTGAAATCATCAGG	TCAAGTATCTAGCCCCAAATCCA
PDGFRA	NM 006206	19	391	GGCAATATTGACCATTCATCATTC	AGGCCAGGAGTAAGACGCAAC
PDGFRB	NM 002609	18	393	AAGAACGTACGTGTGGTGTTGG	CGCTATACTTGCTCCATGCACT
PDGFRB	NM 002609	19	395	AGGAAACAGCCTCTGGTCCTC	GTCAATGCTCAGACAGGGAGAT
PTK7	NM 002821	18	397	CCCAGGAAGGCAGGTACTGTTA	TTTTACAACCACCAAGGGTGTG
PTK7	NM 002821	19	399	TCGTGTGGTTACCTCCAGATTTT	AAATTAGCCAGGGAGTGGAGGT
RET	NM 000323	15	401	CATGCCATGCTATGGCTCAC	AGGCTGAGCGGAGTTCTAATTG
RET	NM 000323	16	403	ATCTCAGCAATCCACAGGAGGT	ATTTGCCTCACGAACACATCAT
ROR1	NM 005012	09 1	405	TGGAAAGTTGTCTATGGCACCTC	ATGGGCAGCAAGGACTTACTCT
ROR1	NM 005012	09 2	407	CACCCCAATATTGTCTGCCTTC	GGCTCGGGAACATGTAATTAGG
ROR1	NM 005012	09 3	409	CCATCATGTATGGCAAATTCTCTT	TGGCGTCTCCTAGTAAAGATGCT
ROR1	NM 005012	09 4	411	GCCAGATTGCTGGTTTCATTG	GGCTAAAACACAAAGCACCATT
RYK	NM 002958	13	413	GGGAAGTCATCCACAAAGACCT	GGTCTGGGTCACAGCTCCTC
TEK	NM 000459	18	415	TTCTTCTGCCAAGATGTGGTGT	TGCAGATGCTGCAATCATGTTA
TEK	NM 000459	19	417	TGGACCCCGAAAGATAAATAGG	TTCTGCACTCCTCTGGAAACTG
TIE	NM 005424	18	419	GGGTGAGAGCCAACACTGATCT	CTGTGCCCTCTCATCTCACACT
TIE	NM 005424	19	421	AGAACCTAGCCTCCAAGATTGC	ACACCTTCCAAGACTCCTTCCA
TYRO3	NM 006293	18	423,424	GACTCGAGGGTGGGAGACAG	GCTGTCACTAGGTGTCCTGAGC

Table S2: *EGFR* mutation status in untreated lung cancer

Sample	Histology	Source	Gender	Exon	Sequence alteration	SEQ ID NO	Nucleotide	Amino acid
S0514	adenocarcinoma	U.S.	F	18	Substitution	425	2155G>A	G719S
S0377	adenocarcinoma	Japan	F	18	Substitution	426	2155G>A	G719S
S0418	adenocarcinoma	Japan	F	19	Del-1a	427	2235_2249delGGAATTAAGAGAAGC	E746_A750del
S0363	large cell ca.	Japan	F	19	Del-1a	428	2235_2249delGGAATTAAGAGAAGC	E746_A750del
S0380	adenocarcinoma	Japan	M	19	Del-1a	429	2235_2249delGGAATTAAGAGAAGC	E746_A750del
S0399	adenocarcinoma	Japan	F	19	Del-1a	430	2235_2249delGGAATTAAGAGAAGC	E746_A750del
S0353	adenocarcinoma	Japan	F	19	Del-1a	431	2235_2249delGGAATTAAGAGAAGC	E746_A750del
S0385	adenocarcinoma	Japan	M	19	Del-1a	432	2235_2249delGGAATTAAGAGAAGC	E746_A750del
S0301	adenocarcinoma	Japan	M	19	Del-1a	433	2235_2249delGGAATTAAGAGAAGC	E746_A750del
S0412	adenocarcinoma	Japan	M	19	Del-1b	434	2236_2250delGAATTAAGAGAAGCA	E746_A750del
S0335	adenocarcinoma	Japan	M	19	Del-1b	435	2236_2250delGAATTAAGAGAAGCA	E746_A750del
S0405	adenocarcinoma	Japan	F	19	Del-1b	436	2236_2250delGAATTAAGAGAAGCA	E746_A750del
S0439	adenocarcinoma	Japan	M	19	Del-2	437	2254_2277delTCTCCGAAAGCCAACAAGGAAATC	S752_I759del
S0361	adenocarcinoma	Japan	F	21	Substitution	438	2573T>G	L858R
S0388	adenocarcinoma	Japan	F	21	Substitution	439	2573T>G	L858R
S0389	adenocarcinoma	Japan	F	21	Substitution	440	2573T>G	L858R

Table S3A: *EGFR* mutation status in gefitinib-treated lung cancer

Gefitinib sensitivity	Sample	Histology	Source	Gender	Exons	Sequence alteration	SEQ ID NO	Nucleotide	Amino acid
Y	IR1T	adenocarcinoma	U.S.	M	19	Del-3	441	2239_2247delTTAAGAGAA, 2248G>C	L747_E749del, A750P
Y	P003	adenocarcinoma	U.S.	M	19	Del-3	442	2239_2247delTTAAGAGAA, 2248G>C	L747_E749del, A750P
Y	IR4T	bronchioloalveolar carcinoma	U.S.	F	19	Del-4	443	2240_2257delTAAGAGAAGCAACATCTC	L747_S752del, P753S
Y	IR2T	adenocarcinoma	U.S.	F	19	Del-5	444	2238_2255delATTAAGAGAAGCAACATC, 2237A>T	L747_S752del, E746V
Y	IR3T	adenocarcinoma	U.S.	F	21	Substitution	445	2573T>G	L858R
Y *in vitro*	IRG H3255	adenocarcinoma	U.S.	F	21	Substitution	446	2573T>G	L858R
N	IR5	adenocarcinoma	U.S.	F	18-24	None detected		n/a	n/a
N	IR6	adenocarcinoma	U.S.	M	18-24	None detected		n/a	n/a
N	IR8	adenocarcinoma	U.S.	F	18-24	None detected		n/a	n/a
N	IR9	NSCLC	U.S.	F	18-24	None detected		n/a	n/a

Table S3B: *EGFR* mutations not shown in Table 2, Table S2, or Table S3A.

Sample	**Tissue**	**Exon**	**Sequence alteration**	**Nucleotide**	**Amino acid**
Tar4T	Lung adenocarcinoma	19	Deletion	2239-2250delTTAAGAGAAGCA; 2251A>C	L747_A750del; T751T
AD355	Lung adenocarcinoma	19	Deletion	2240-2254delTAAGAGAAGCA	L747_T751del
IR TT	Lung adenocarcinoma	19	Deletion	2257-2271delCCGAAAGCCAACAAG	P753_K757del
AD240	Lung adenocarcinoma	20	Insertion	2309-2310insCAACCCGG	D770_N771ins NPG
AD261	Lung adenocarcinoma	20	Insertion	2311-2312insGCGTGGACA	D770_N771ins SVD
	Lung adenocarcinoma	20	Insertion	2316-2317insGGT	P772_H773ins V
AD356	Lung adenocarcinoma	20	Substitution	2334-2335GG>AA	G779S
SP02-23	Acute myeloid leukemia	21	Substitution	2570G>T	G857V
SP08-94	Glioma	21	Substitution	2582T>A	L861Q
SP06-45	Sarcoma	21	Substitution	2648T>C	L883S
AD241	Colon adenocarcinoma	22	Substitution	2686G>T	D896Y

Table S3C Position of BCR-ABL mutants resistant to imatinib and analogous positions in EGFR

Abl1 residue subject to resistance mutation	Analogous EGFR residue	Identical/similar/non-conserved
Met-244	Lys-714	Non-conserved
Leu-248	Leu-718	Identical
Gly-250	Ser-720	Non-conserved
Gln-252	Ala-722	Non-conserved
Tyr-253	Phe-723	Similar
Glu-255	Thr-725	Non-conserved
Asp-276	Ala-750	Non-conservd
Thr-315	Thr-790	Identical
Phe-317	Leu-792	Similar
Met-351	Met-825	Identical
Glu-355	Glu-829	Identical
Phe-359	Leu-833	Similar
His-396	His-870	Identical
Ser-417	Thr-892	Similar
Phe-486	Phe-961	Identical

Table S4: Primers used for cDNA sequencing

Primer name	SEQ ID NO	Primer sequence 5' to 3'
cDNA_EGFR_aF	447	TGTAAAACGACGGCCAGTCGCCCAGACCGGACGACA
cDNA_EGFR_aR	448	CAGGAAACAGCTATGACCAGGGCAATGAGGACATAACCA
cDNA_EGFR_bF	449	TGTAAAACGACGGCCAGTGGTGGTCCTTGGGAATTTGG
cDNA_EGFR_bR	450	CAGGAAACAGCTATGACCCCATCGACATGTTGCTGAGAAA
cDNA_EGFR_cF	451	TGTAAAACGACGGCCAGTGAAGGAGCTGCCCATGAGAA
cDNA_EGFR_cR	452	CAGGAAACAGCTATGACCCGTGGCTTCGTCTCGGAATT
cDNA_EGFR_dF	453	TGTAAAACGACGGCCAGTGAAACTGACCAAAATCATCTGT
cDNA_EGFR_dR	454	CAGGAAACAGCTATGACCTACCTATTCCGTTACACACTTT
cDNA_EGFR_eF	455	TGTAAAACGACGGCCAGTCCGTAATTATGTGGTGACAGAT
cDNA_EGFR_cR	456	CAGGAAACAGCTATGACCGCGTATGATTTCTAGGTTCTCA
cDNA_EGFR_fF	457	TGTAAAACGACGGCCAGTCTGAAAACCGTAAAGGAAATCAC
cDNA_EGFR_fR	458	CAGGAAACAGCTATGACCCCTGCCTCGGCTGACATTC
cDNA_EGFR_gF	459	TGTAAAACGACGGCCAGTTAAGCAACAGAGGTGAAAACAG
cDNA_EGFR_gR	460	CAGGAAACAGCTATGACCGGTGTTGTTTTCTCCCATGACT
cDNA_EGFR_hF	461	TGTAAAACGACGGCCAGTGGACCAGACAACTGTATCCA
cDNA_EGFR_hR	462	CAGGAAACAGCTATGACCTTCCTTCAAGATCCTCAAGAGA
cDNA_EGFR_iF	463	TGTAAAACGACGGCCAGTGATCGGCCTCTTCATGCGAA
cDNA_EGFR_iR	464	CAGGAAACAGCTATGACCACGGTGGAGGTGAGGCAGAT
cDNA_EGFR_jF	465	TGTAAAACGACGGCCAGTCGAAAGCCAACAAGGAAATCC
cDNA_EGFR_jR	466	CAGGAAACAGCTATGACCATTCCAATGCCATCCACTTGAT
cDNA_EGFR_kF	467	TGTAAAACGACGGCCAGTAACACCGCAGCATGTCAAGAT
cDNA_EGFR_kR	468	CAGGAAACAGCTATGACCCTCGGGCCATTTTGGAGAATT
cDNA_EGFR_lF	469	TGTAAAACGACGGCCAGTTCAGCCACCCATATGTACCAT
cDNA_EGFR_lR	470	CAGGAAACAGCTATGACCGCTTTGCAGCCCATTTCTATC
cDNA_EGFR_mF	471	TGTAAAACGACGGCCAGTACAGCAGGGCTTCTTCAGCA
cDNA_EGFR_mR	472	CAGGAAACAGCTATGACCTGACACAGGTGGGCTGGACA
cDNA_EGFR_nF	473	TGTAAAACGACGGCCAGTGAATCCTGTCTATCACAATCAG
cDNA_EGFR_nR	474	CAGGAAACAGCTATGACCGGTATCGAAAGAGTCTGGATTT
cDNA_EGFR_oF	475	TGTAAAACGACGGCCAGTGCTCCACAGCTGAAAATGCA
cDNA_EGFR_oR	476	CAGGAAACAGCTATGACCACGTTGCAAAACCAGTCTGTG

THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. A method for determining the likelihood of effectiveness of an epidermal growth factor receptor (EGFR) targeting treatment in a human patient affected with or at risk for developing cancer comprising: detecting the presence or absence of at least one nucleic acid variance in the kinase domain of the erbB 1 gene of said patient relative to the wildtype erbB 1 gene, wherein the presence of the at least one nucleic acid variance indicates that the EGFR targeting treatment is likely to be effective.

2. The method of claim 1, wherein the nucleic acid variance increases kinase activity.

3. The method of claim 1, wherein the erbB 1 gene is obtained from a biological sample from said patient.

4. The method of claim 1, wherein the variance in the kinase domain of the erbB 1 gene effects the conformational structure of the ATP-binding pocket.

5. The method of claim1, wherein the variance in the kinase domain of erbB 1 is in an exon of the erbB 1 gene selected from the group consisting of exon 18, 19, 20 or 21.

6. The method of claim 5, wherein the variance is in exon 18, 19 or 21.

7. The method of claim 1, wherein the variance in the kinase domain of the erbB 1 gene is an in frame deletion, substitution, or insertion.

8. The method of claim 7, wherein the in frame deletion is in exon 19 of the erbB 1 gene.

9. The method of claim 8, wherein the in frame deletion in exon 19 of the erbB 1 gene comprises a deletion of at least amino acids leucine, arginine, glutamic acid and alanine, at codons 747, 748, 749, and 750 of SEQ. ID. NO. 512.

10. The method of claim 8, wherein the in frame deletion in exon 19 of the erbB 1 gene comprises a deletion of at least amino acids leucine, arginine, and glutamic acid at codon 747, 748 and 749 of SEQ. ID. NO. 512.

11. The method of claim 8, wherein the in frame deletion comprises nucleotides selected from the group consisting of 2235 to 2249, 2240 to 2251, and 2240 to 2257 of SEQ ID NO: 511.

12. The method of claim 7, wherein the substitution is in exon 21 of the erbB 1 gene.

13. The method of claim 12, wherein the substitution in exon 21 comprises at least one codon.

14. The method of claim 12, wherein the substitution in exon 21 comprises a substitution from the group consisting of a guanine for a thymine at nucleotide 2573 of SEQ ID NO:511, and an adenine for a thymine at nucleotide 2582 of SEQ ID NO:511.

15. The method of claim 7, wherein the substitution is in exon 18 of the erbB 1 gene.

16. The method of claim 15, wherein the substitution in exon 18 is a thymine for a guanine or an adenine for a guanine at nucleotide 2155 of SEQ ID NO: 511.

17. The method of any one of claims 1 to 16, wherein the detection of the presence or absence of said at least one variance comprises amplifying a segment of nucleic acid.

18. The method of claim 17, wherein the segment to be amplified is 1000 nucleotides in length or less.

19. The method of claim 17, wherein the segment to be amplified includes a plurality of variances.

20. The method of any one of claims 1 to 16, wherein the detection of the presence or absence of said at least one variance comprises contacting the erbB 1 nucleic acid with at least one nucleic acid probe, wherein said at least one probe preferentially hybridizes with a nucleic acid sequence comprising said variance under selective hybridization conditions.

21. The method of any one of claims 1 to 16, wherein the detection of the presence or absence of said at least one variance comprises sequencing at least one nucleic acid sequence.

22. The method of any one of claims 1 to 16, wherein the detection of the presence or absence of said at least one variance comprises mass spectrometric determination of at least one nucleic acid sequence.

23. The method of any one of claims 1 to 16, wherein the detection of the presence or absence of said at least one variance comprises performing a polymerase chain reaction (PCR) to amplify nucleic acid comprising the erbBl coding sequence, and determining nucleotide sequence of the amplified nucleic acid.

24. The method of claim 23, wherein determining the nucleotide sequence of the amplified nucleic acid comprises sequencing at least one nucleic acid segment.

25. The method of claim 23, wherein determining the nucleotide sequence of the amplified nucleic acid comprises running the amplified nucleic acid segment on a gel and determining the segment's size.

26. The method of any one of claims 1 to 16, wherein the detection of the presence or absence of said at least one variance comprises determining the haplotype of a plurality of variances in a gene.

27. The method of any one of the preceding claims , wherein the EGFR targeting treatment is a tyrosine kinase inhibitor.

28. The method of claim 27, wherein the tyrosine kinase inhibitor is an anilinoquinazoline.

29. The method of claim 28, wherein the anilinoquinazolineis a synthetic anilinoquinazoline.

30. The method of claim 29, wherein the synthetic anilinoquinazoline is selected from the group consisting of gefitinib and erlotinib.

31. A method for determining the likelihood of effectiveness of an epidermal growth factor receptor (EGFR) targeting treatment in a human patient affected with or at risk for developing cancer comprising:

  a. detecting the presence or absence of at least one nucleic acid variance in exon 18, 19, 20, or 21 by performing a polymerase chain reaction (PCR) to amplify a portion of exon 18, 19, 20, or 21 ; and
  b. determining the nucleotide sequence of the amplified nucleic acid by sequencing at least one portion of the amplified exon 18,19, 20, or 21, wherein the presence of at least one nucleotide variance in exon 18,19, 20 or 21

compared to a wildtype erbB 1 control indicates that the EGFR targeting treatment is likely to be effective.

32. A method of treating a patient affected with or at risk for developing cancer, comprising detecting the presence or absence of at least one nucleic acid variance in the kinase domain of the erbB 1 gene of the patient, wherein the patient is administered an EGFR targeting treatment if the presence of the said at least one nucleic acid variance is detected.

33. A method of treating a patient affected with or at risk for developing cancer, comprising:

a. detecting the presence or absence of at least one nucleic acid variance in exon 18, 19, 20, or 21 by performing a polymerase chain reaction (PCR) to amplify a portion of exon 18, 19, 20, or 21,

b. determining the nucleotide sequence of the amplified nucleic acid by sequencing at least one portion of the amplified exon 18, 19, 20 or 21 ; and

c. and administering an EGFR targeting treatment to the patient if the presence of the said at least one nucleic acid variance is detected.

34. The method of claim 32 or 33, wherein the EGFR targeting treatment is a tyrosine kinase inhibitor.

35. The method of claim 34, wherein the tyrosine kinase inhibitor is an anilinoquinazoline.

36. The method of claim 35, wherein the anilinoquinazoline is a synthetic anilinoquinazoline.

37. The method of claim 36, wherein the synthetic anilinoquinazoline is selected from the group consisting of gefitinib and erlotinib.

38. The method of claim 32 and 33, wherein the erbB 1 gene is obtained from a biological sample from said patient.

39. The method of claim 32 and 33, wherein said cancer is selected from the group consisting of gastrointestinal cancer, prostate cancer, ovarian cancer, breast cancer, head and neck cancer, lung cancer, non-small cell lung cancer, cancer of the nervous

system, kidney cancer, retina cancer, skin cancer, liver cancer, pancreatic cancer, genital-urinary cancer and bladder cancer.

40. The method of claim 39, wherein said cancer is non-small cell lung cancer.

41. The method of claim 32 and 33, wherein the nucleic acid variance increases kinase activity.

42. The method of claim 32 and 33, wherein the variance in the kinase domain of the erbB 1 gene effects the conformation of the ATP-binding pocket.

43. The method of claim 32 and 33, wherein the variance in the kinase domain of erbB 1 is in an exon of the erbB 1 gene selected from the group consisting of exon 18, 19, 20 or 21.

44. The method of claim 43, wherein the variance is in exon 18, 19 or 21.

45. The method of claim 32 and 33, wherein the variance in the kinase domain of the erbB 1 gene is an in frame deletion or a substitution.

46. The method of claim 45, wherein the in frame deletion is in exon 19 of the erbB 1 gene.

47. The method of claim 46, wherein the in frame deletion in exon 19 of the erbB 1 gene comprises a deletion of at least amino acids leucine, arginine, glutamic acid and alanine, at codons 747, 748, 749, and 750 of SEQ. ID. NO. 512.

48. The method of claim 46, wherein the in frame deletion in exon 19 of the erbB 1 gene comprises a deletion of at least amino acids leucine, arginine, and glutamine acid at codon 747, 748 and 749 of SEQ. ID. NO. 512.

49. The method of claim 47, wherein the in frame deletion comprises nucleotides selected from the group consisting of 2235 to 2249, 2240 to 2251, and 2240 to 2257 of SEQ ID NO: 511.

50. The method of claim 45, wherein the substitution is in exon 21 of the erbB 1 gene.

51. The method of claim 50, wherein the substitution in exon 21 comprises a substitution from the group consisting of a guanine for a thymine at nucleotide 2573 of SEQ ID NO:511, and an adenine for a thymine at nucleotide 2582 of SEQ ID NO: 511.

52. The method of claim 50, wherein the substitution in exon 21 comprises at least one codon.

53. The method of claim 45, wherein the substitution is in exon 18 of the erbB 1 gene.

54. The method of claim 53, wherein the substitution in exon 18 is a thymine for a guanine at nucleotide 2155 of SEQ ID NO: 511.

55. A kit when used in the method of claim 17 or claim 31 comprising:
a. at least one degenerate primer pair designed to anneal to nucleic acid regions bordering or within exon 18, 19, 20 or 21 of the EGFR kinase domain;
b. products and reagents required to carry out PCR amplification; and
c. instructions.

56. A kit according to claim 55, wherein the primer comprises the sequence primers selected from the group consisting of SEQ ID NOS: 505-508, and SEQ ID NOS. 646-673 with SEQ ID NO. 645 on the 5' end of all forward primers and SEQ ID NO. 674 on the 5' end of all reverse primers.

57. A kit when used in the method of claim 1 comprising:
a. at least one probe designed to anneal to nucleic acid regions within exons 18, 19, 20 or 21 of the EGFR kinase domain;
b. products and reagents required to carry out the annealing reaction; and
c. instructions.

58. The kit of claim 57 wherein at least one probe is bound to a solid support.

59. A kit when used in the method of claim 1 comprising:

a. at least one probe designed to bind to the ATP-binding pocket of the EGFR kinase domain protein;

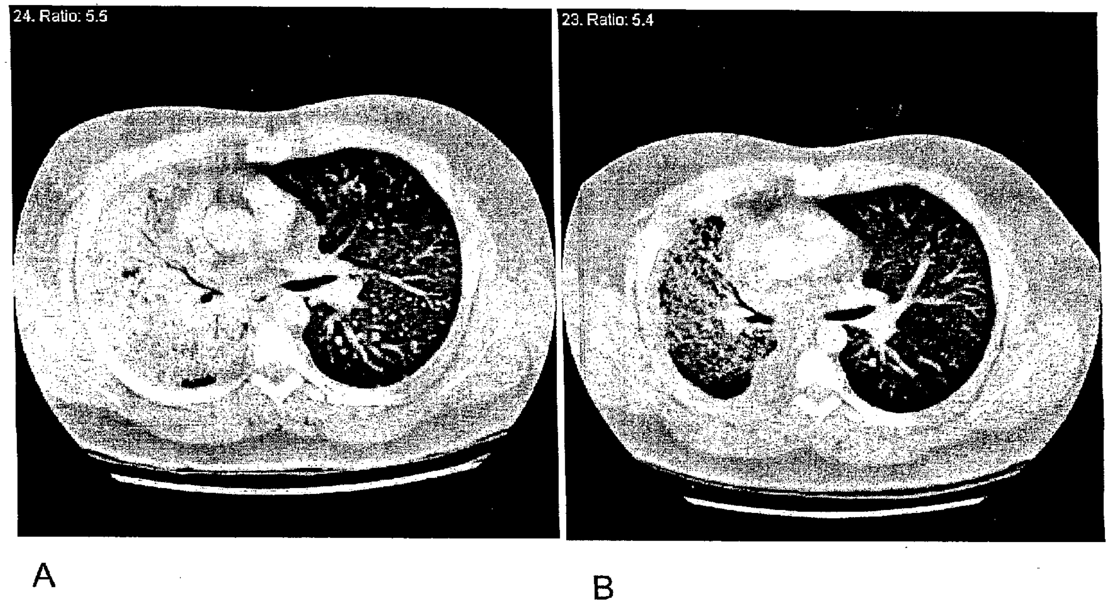
b. the products and reagents required to carry out the binding reaction; and

c. instructions.

60. The kit according to claim 59, wherein the probe is an antibody, antibody fragment, or chimeric antibody.

61. The kit according to claim 60, wherein the probe further comprises a detectable label.

Figure 1 A and B



Figures 2A - 2E

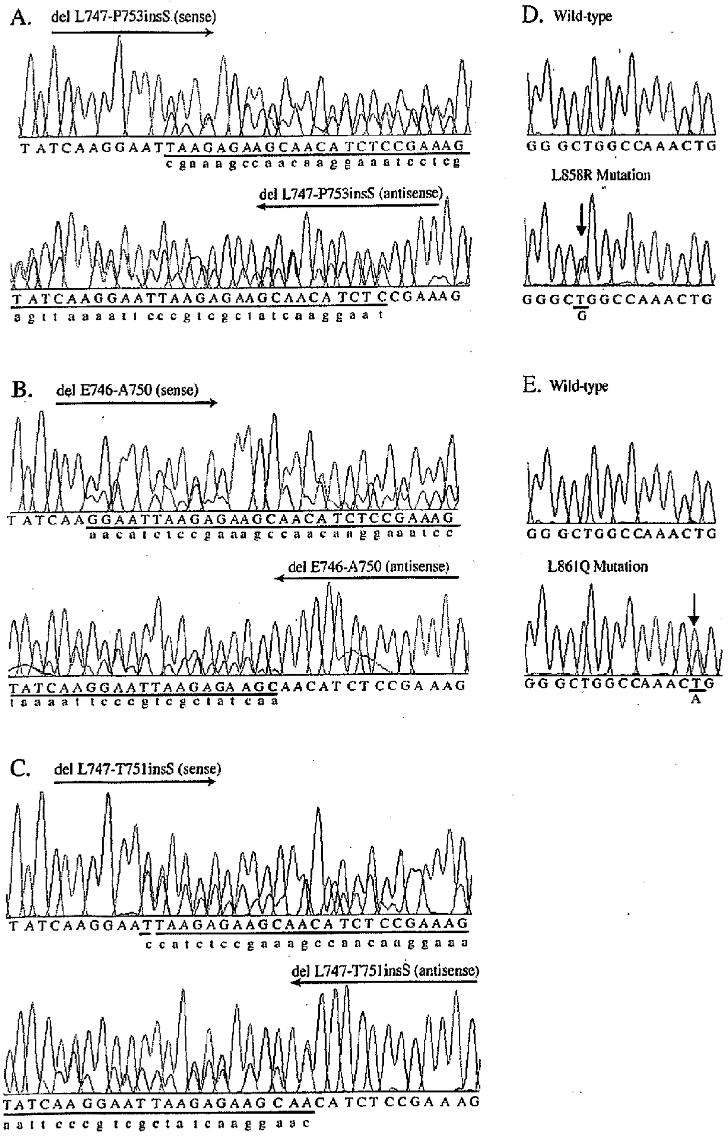
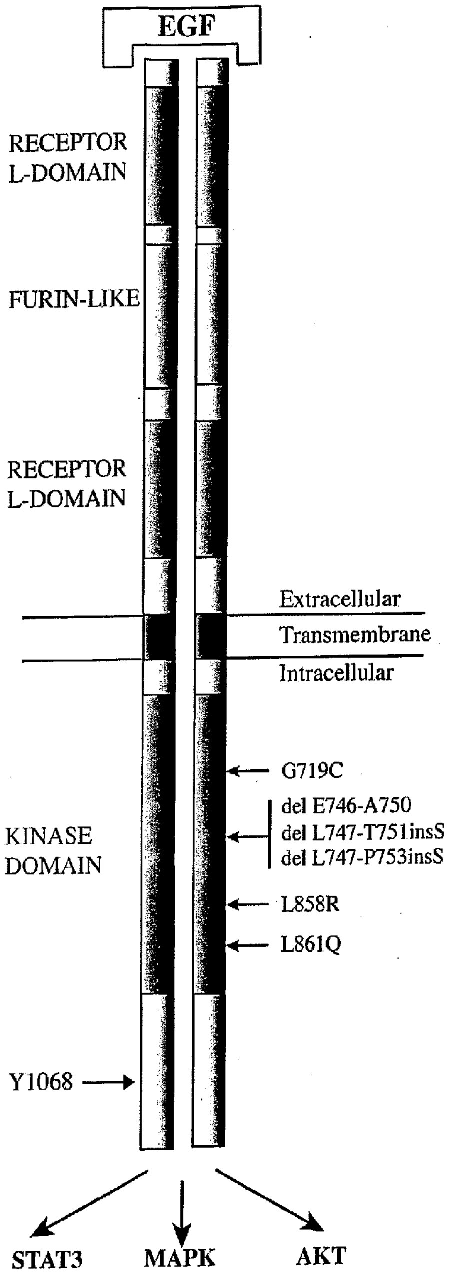
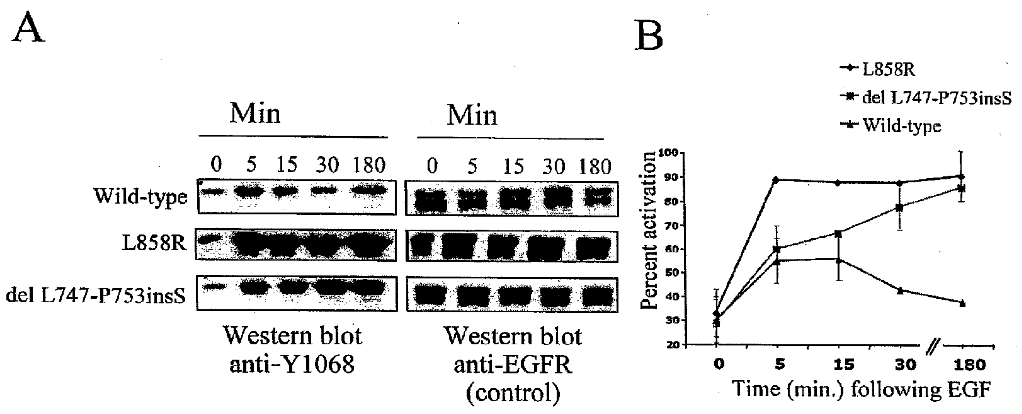


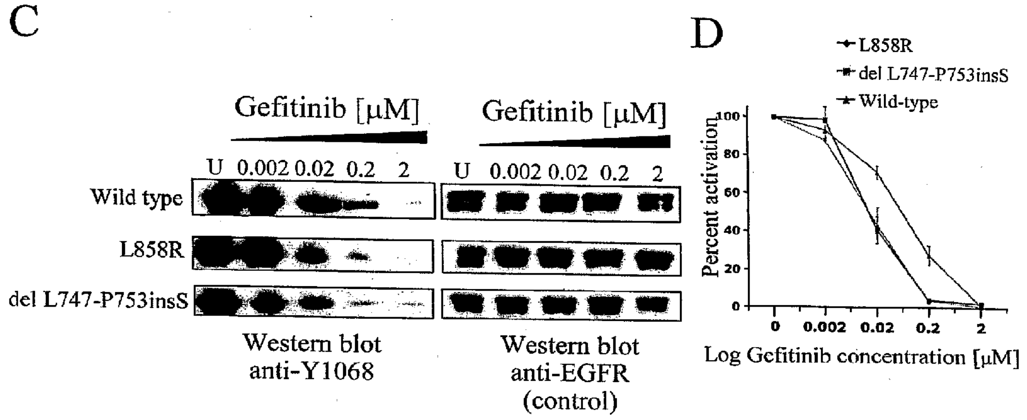
Figure 2F


EGF
RECEPTOR L-DOMAIN
FURIN-LIKE
RECEPTOR L-DOMAIN
Extracellular
Transmembrane
Intracellular
KINASE DOMAIN
G719C
del E746-A750
del L747-T751insS
del L747-P753insS
L858R
L861Q
Y1068
STAT3
MAPK
AKT

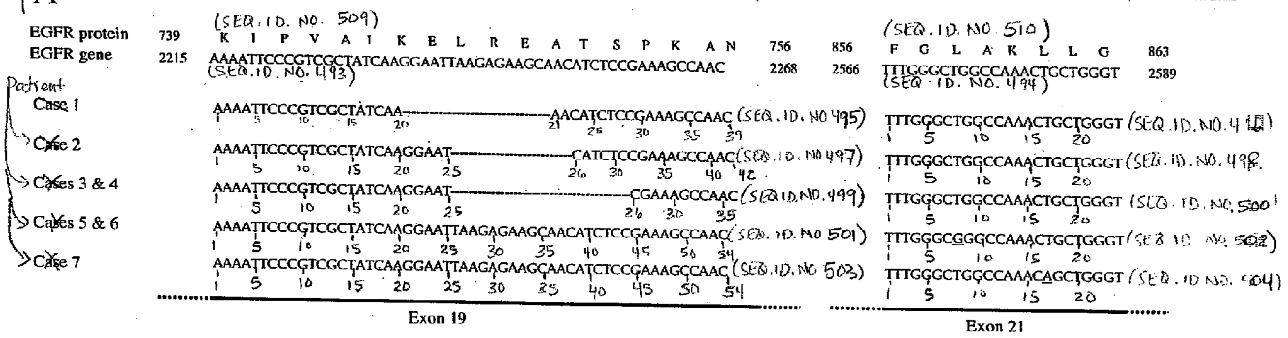
Figures 3A and 3B



Figures 3C and 3D



4A


(SEQ.ID.NO.509)
EGFR protein 739 K I P V A I K E L R E A T S P K A N 756
EGFR gene 2215 AAAATTCCCGTCGCTATCAAGGAATTAAGAGAAGCAACATCTCCGAAAGCCAAC 2268
(SEQ.ID.NO.493)
(SEQ.ID.NO.510)
856 F G L A K L L G 863
2566 TTTGGGCTGGCCAAACTGCTGGGT 2589
(SEQ.ID.NO.494)
Patient
Case 1
Case 2
Cases 3 & 4
Cases 5 & 6
Case 7
AAAATTCCCGTCGCTATCAA——AACATCTCCGAAAGCCAAC (SEQ.ID.NO 495)
1 5 10 15 20 21 25 30 35 39
TTTGGGCTGGCCAAACTGCTGGGT (SEQ.ID.NO.496)
1 5 10 15 20
AAAATTCCCGTCGCTATCAAGGAAT——CATCTCCGAAAGCCAAC (SEQ.ID.NO 497)
1 5 10 15 20 25 26 30 35 40 42
TTTGGGCTGGCCAAACTGCTGGGT (SEQ.ID.NO.498
1 5 10 15 20
AAAATTCCCGTCGCTATCAAGGAAT——CGAAAGCCAAC (SEQ.ID.NO.499)
1 5 10 15 20 25 26 30 35
TTTGGGCTGGCCAAACTGCTGGGT (SEQ.ID.NO.500
1 5 10 15 20
AAAATTCCCGTCGCTATCAAGGAATTAAGAGAAGCAACATCTCCGAAAGCCAAC (SEQ.ID.NO 501)
1 5 10 15 20 25 30 35 40 45 50 54
TTTGGGCGGGCCAAACTGCTGGGT (SEQ.ID.NO.502)
1 5 10 15 20
AAAATTCCCGTCGCTATCAAGGAATTAAGAGAAGCAACATCTCCGAAAGCCAAC (SEQ.ID.NO 503)
1 5 10 15 20 25 30 35 40 45 50 54
TTTGGGCTGGCCAAACAGCTGGGT (SEQ.ID.NO.504)
1 5 10 15 20
Exon 19
Exon 21

Figures 4B-C

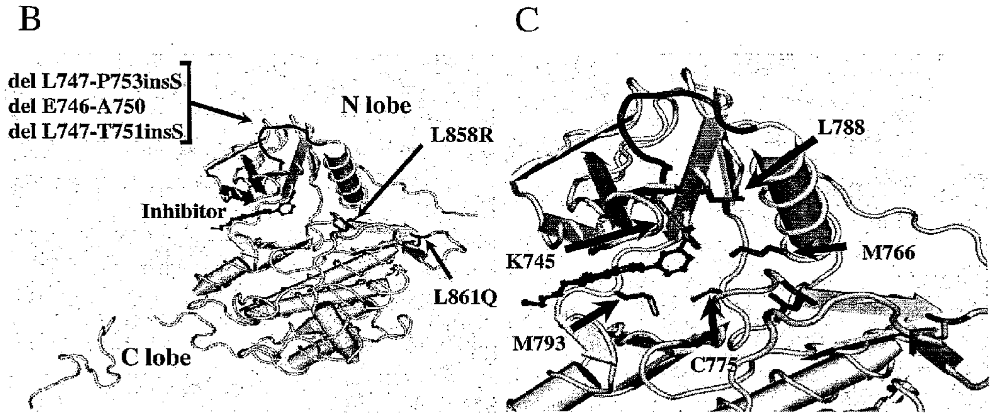

B
del L747-P753insS
del E746-A750
del L747-T751insS
N lobe
L858R
Inhibitor
L861Q
C lobe
C
L788
K745
M766
M793
C775

Figure 5

```
CCCGGCGCAGCGCGGCCGCAGCAGCCTCCGCCCCCCGCACGGTGTGAGCGCCCGACGCGG -185
............................................................

CCGAGGCGGCCGGAGTCCCGAGCTAGCCCCGGCGGCCGCCGCCGCCCAGACCGGACGACA -125
............................................................

GGCCACCTCGTCGGCGTCCGCCCGAGTCCCCGCCTCGCCGCCAACGCCACAACCACCGCG -65
............................................................

CACGGCCCCCTGACTCCGTCCAGTATTGATCGGGAGAGCCGGAGCGAGCTCTTCGGGGAG -5
............................................................

CAGCGATGCGACCCTCCGGGACGGCCGGGGCAGCGCTCCTGGCGCTGCTGGCTGCGCTCT 55
.....-M--R--P--S--G--T--A--G--A--A--L--L--A--L--L--A--A--L-- 18

GCCCGGCGAGTCGGGCTCTGGAGGAAAAGAAAGTTTGCCAAGGCACGAGTAACAAGCTCA 115
C--P--A--S--R--A--L--E--E--K--K--V--C--Q--G--T--S--N--K--L-- 38

CGCAGTTGGGCACTTTTGAAGATCATTTTCTCAGCCTCCAGAGGATGTTCAATAACTGTG 175
T--Q--L--G--T--F--E--D--H--F--L--S--L--Q--R--M--F--N--N--C-- 58

AGGTGGTCCTTGGGAATTTGGAAATTACCTATGTGCAGAGGAATTATGATCTTTCCTTCT 235
E--V--V--L--G--N--L--E--I--T--Y--V--Q--R--N--Y--D--L--S--F-- 78

TAAAGACCATCCAGGAGGTGGCTGGTTATGTCCTCATTGCCCTCAACACAGTGGAGCGAA 295
L--K--T--I--Q--E--V--A--G--Y--V--L--I--A--L--N--T--V--E--R-- 98

TTCCTTTGGAAAACCTGCAGATCATCAGAGGAAATATGTACTACGAAAATTCCTATGCCT 355
I--P--L--E--N--L--Q--I--I--R--G--N--M--Y--Y--E--N--S--Y--A-- 118

TAGCAGTCTTATCTAACTATGATGCAAATAAAACCGGACTGAAGGAGCTGCCCATGAGAA 415
L--A--V--L--S--N--Y--D--A--N--K--T--G--L--K--E--L--P--M--R-- 138

ATTTACAGGAAATCCTGCATGGCGCCGTGCGGTTCAGCAACAACCCTGCCCTGTGCAACG 475
N--L--Q--E--I--L--H--G--A--V--R--F--S--N--N--P--A--L--C--N-- 158

TGGAGAGCATCCAGTGGCGGGACATAGTCAGCAGTGACTTTCTCAGCAACATGTCGATGG 535
V--E--S--I--Q--W--R--D--I--V--S--S--D--F--L--S--N--M--S--M-- 178

ACTTCCAGAACCACCTGGGCAGCTGCCAAAAGTGTGATCCAAGCTGTCCCAATGGGAGCT 595
D--F--Q--N--H--L--G--S--C--Q--K--C--D--P--S--C--P--N--G--S-- 198

GCTGGGGTGCAGGAGAGGAGAACTGCCAGAAACTGACCAAAATCATCTGTGCCCAGCAGT 655
C--W--G--A--G--E--E--N--C--Q--K--L--T--K--I--I--C--A--Q--Q-- 218

GCTCCGGGCGCTGCCGTGGCAAGTCCCCCAGTGACTGCTGCCACAACCAGTGTGCTGCAG 715
C--S--G--R--C--R--G--K--S--P--S--D--C--C--H--N--Q--C--A--A-- 238

GCTGCACAGGCCCCCGGGAGAGCGACTGCCTGGTCTGCCGCAAATTCCGAGACGAAGCCA 775
G--C--T--G--P--R--E--S--D--C--L--V--C--R--K--F--R--D--E--A-- 258

CGTGCAAGGACACCTGCCCCCCACTCATGCTCTACAACCCCACCACGTACCAGATGGATG 835
T--C--K--D--T--C--P--P--L--M--L--Y--N--P--T--T--Y--Q--M--D-- 278

TGAACCCCGAGGGCAAATACAGCTTTGGTGCCACCTGCGTGAAGAAGTGTCCCCGTAATT 895
V--N--P--E--G--K--Y--S--F--G--A--T--C--V--K--K--C--P--R--N-- 298
```

Figure 5 (cont.)

```
ATGTGGTGACAGATCACGGCTCGTGCGTCCGAGCCTGTGGGGCCGACAGCTATGAGATGG 955
Y--V--V--T--D--H--G--S--C--V--R--A--C--G--A--D--S--Y--E--M-- 318

AGGAAGACGGCGTCCGCAAGTGTAAGAAGTGCGAAGGGCCTTGCCGCAAAGTGTGTAACG 1015
E--E--D--G--V--R--K--C--K--K--C--E--G--P--C--R--K--V--C--N-- 338

GAATAGGTATTGGTGAATTTAAAGACTCACTCTCCATAAATGCTACGAATATTAAACACT 1075
G--I--G--I--G--E--F--K--D--S--L--S--I--N--A--T--N--I--K--H-- 358

TCAAAAACTGCACCTCCATCAGTGGCGATCTCCACATCCTGCCGGTGGCATTTAGGGGTG 1135
F--K--N--C--T--S--I--S--G--D--L--H--I--L--P--V--A--F--R--G-- 378

ACTCCTTCACACATACTCCTCCTCTGGATCCACAGGAACTGGATATTCTGAAAACCGTAA 1195
D--S--F--T--H--T--P--P--L--D--P--Q--E--L--D--I--L--K--T--V-- 398

AGGAAATCACAGGGTTTTTGCTGATTCAGGCTTGGCCTGAAAACAGGACGGACCTCCATG 1255
K--E--I--T--G--F--L--L--I--Q--A--W--P--E--N--R--T--D--L--H-- 418

CCTTTGAGAACCTAGAAATCATACGCGGCAGGACCAAGCAACATGGTCAGTTTTCTCTTG 1315
A--F--E--N--L--E--I--I--R--G--R--T--K--Q--H--G--Q--F--S--L-- 438

CAGTCGTCAGCCTGAACATAACATCCTTGGGATTACGCTCCCTCAAGGAGATAAGTGATG 1375
A--V--V--S--L--N--I--T--S--L--G--L--R--S--L--K--E--I--S--D-- 458

GAGATGTGATAATTTCAGGAAACAAAAATTTGTGCTATGCAAATACAATAAACTGGAAAA 1435
G--D--V--I--I--S--G--N--K--N--L--C--Y--A--N--T--I--N--W--K-- 478

AACTGTTTGGGACCTCCGGTCAGAAAACCAAAATTATAAGCAACAGAGGTGAAAACAGCT 1495
K--L--F--G--T--S--G--Q--K--T--K--I--I--S--N--R--G--E--N--S-- 498

GCAAGGCCACAGGCCAGGTCTGCCATGCCTTGTGCTCCCCCGAGGGCTGCTGGGGCCCGG 1555
C--K--A--T--G--Q--V--C--H--A--L--C--S--P--E--G--C--W--G--P-- 518

AGCCCAGGGACTGCGTCTCTTGCCGGAATGTCAGCCGAGGCAGGGAATGCGTGGACAAGT 1615
E--P--R--D--C--V--S--C--R--N--V--S--R--G--R--E--C--V--D--K-- 538

GCAACCTTCTGGAGGGTGAGCCAAGGGAGTTTGTGGAGAACTCTGAGTGCATACAGTGCC 1675
C--N--L--L--E--G--E--P--R--E--F--V--E--N--S--E--C--I--Q--C-- 558

ACCCAGAGTGCCTGCCTCAGGCCATGAACATCACCTGCACAGGACGGGGACCAGACAACT 1735
H--P--E--C--L--P--Q--A--M--N--I--T--C--T--G--R--G--P--D--N-- 578

GTATCCAGTGTGCCCACTACATTGACGGCCCCCACTGCGTCAAGACCTGCCCGGCAGGAG 1795
C--I--Q--C--A--H--Y--I--D--G--P--H--C--V--K--T--C--P--A--G-- 598

TCATGGGAGAAAACAACACCCTGGTCTGGAAGTACGCAGACGCCGGCCATGTGTGCCACC 1855
V--M--G--E--N--N--T--L--V--W--K--Y--A--D--A--G--H--V--C--H-- 618

TGTGCCATCCAAACTGCACCTACGGATGCACTGGGCCAGGTCTTGAAGGCTGTCCAACGA 1915
L--C--H--P--N--C--T--Y--G--C--T--G--P--G--L--E--G--C--P--T-- 638

ATGGGCCTAAGATCCCGTCCATCGCCACTGGGATGGTGGGGGCCCTCCTCTTGCTGCTGG 1975
N--G--P--K--I--P--S--I--A--T--G--M--V--G--A--L--L--L--L--L-- 658

TGGTGGCCCTGGGGATCGGCCTCTTCATGCGAAGGCGCCACATCGTTCGGAAGCGCACGC 2035
V--V--A--L--G--I--G--L--F--M--R--R--R--H--I--V--R--K--R--T-- 678
```

9/19

Figure 5 (cont.)

```
TGCGGAGGCTGCTGCAGGAGAGGGAGCTTGTGGAGCCTCTTACACCCAGTGGAGAAGCTC 2095
L--R--R--L--L--Q--E--R--E--L--V--E--P--L--T--P--S--G--E--A-- 698

CCAACCAAGCTCTCTTGAGGATCTTGAAGGAAACTGAATTCAAAAAGATCAAAGTGCTGG 2155
P--N--Q--A--L--L--R--I--L--K--E--T--E--F--K--K--I--K--V--L-- 718

GCTCCGGTGCGTTCGGCACGGTGTATAAGGGACTCTGGATCCCAGAAGGTGAGAAAGTTA 2215
G--S--G--A--F--G--T--V--Y--K--G--L--W--I--P--E--G--E--K--V-- 738

AAATTCCCGTCGCTATCAAGGAATTAAGAGAAGCAACATCTCCGAAAGCCAACAAGGAAA 2275
K--I--P--V--A--I--K--E--L--R--E--A--T--S--P--K--A--N--K--E-- 758

TCCTCGATGAAGCCTACGTGATGGCCAGCGTGGACAACCCCCACGTGTGCCGCCTGCTGG 2335
I--L--D--E--A--Y--V--M--A--S--V--D--N--P--H--V--C--R--L--L-- 778

GCATCTGCCTCACCTCCACCGTGCAGCTCATCACGCAGCTCATGCCCTTCGGCTGCCTCC 2395
G--I--C--L--T--S--T--V--Q--L--I--T--Q--L--M--P--F--G--C--L-- 798

TGGACTATGTCCGGGAACACAAAGACAATATTGGCTCCCAGTACCTGCTCAACTGGTGTG 2455
L--D--Y--V--R--E--H--K--D--N--I--G--S--Q--Y--L--L--N--W--C-- 818

TGCAGATCGCAAAGGGCATGAACTACTTGGAGGACCGTCGCTTGGTGCACCGCGACCTGG 2515
V--Q--I--A--K--G--M--N--Y--L--E--D--R--R--L--V--H--R--D--L-- 838

CAGCCAGGAACGTACTGGTGAAAACACCGCAGCATGTCAAGATCACAGATTTTGGGCTGG 2575
A--A--R--N--V--L--V--K--T--P--Q--H--V--K--I--T--D--F--G--L-- 858

CCAAACTGCTGGGTGCGGAAGAGAAAGAATACCATGCAGAAGGAGGCAAAGTGCCTATCA 2635
A--K--L--L--G--A--E--E--K--E--Y--H--A--E--G--G--K--V--P--I-- 878

AGTGGATGGCATTGGAATCAATTTTACACAGAATCTATACCCACCAGAGTGATGTCTGGA 2695
K--W--M--A--L--E--S--I--L--H--R--I--Y--T--H--Q--S--D--V--W-- 898

GCTACGGGGTGACTGTTTGGGAGTTGATGACCTTTGGATCCAAGCCATATGACGGAATCC 2755
S--Y--G--V--T--V--W--E--L--M--T--F--G--S--K--P--Y--D--G--I-- 918

CTGCCAGCGAGATCTCCTCCATCCTGGAGAAAGGAGAACGCCTCCCTCAGCCACCCATAT 2815
P--A--S--E--I--S--S--I--L--E--K--G--E--R--L--P--Q--P--P--I-- 938

GTACCATCGATGTCTACATGATCATGGTCAAGTGCTGGATGATAGACGCAGATAGTCGCC 2875
C--T--I--D--V--Y--M--I--M--V--K--C--W--M--I--D--A--D--S--R-- 958

CAAAGTTCCGTGAGTTGATCATCGAATTCTCCAAAATGGCCCGAGACCCCCAGCGCTACC 2935
P--K--F--R--E--L--I--I--E--F--S--K--M--A--R--D--P--Q--R--Y-- 978

TTGTCATTCAGGGGGATGAAAGAATGCATTTGCCAAGTCCTACAGACTCCAACTTCTACC 2995
L--V--I--Q--G--D--E--R--M--H--L--P--S--P--T--D--S--N--F--Y-- 998

GTGCCCTGATGGATGAAGAAGACATGGACGACGTGGTGGATGCCGACGAGTACCTCATCC 3055
R--A--L--M--D--E--E--D--M--D--D--V--V--D--A--D--E--Y--L--I-- 1018

CACAGCAGGGCTTCTTCAGCAGCCCCTCCACGTCACGGACTCCCCTCCTGAGCTCTCTGA 3115
P--Q--Q--G--F--F--S--S--P--S--T--S--R--T--P--L--L--S--S--L-- 1038

GTGCAACCAGCAACAATTCCACCGTGGCTTGCATTGATAGAAATGGGCTGCAAAGCTGTC 3175
S--A--T--S--N--N--S--T--V--A--C--I--D--R--N--G--L--Q--S--C-- 1058
```

Figure 5 (cont.)

```
CCATCAAGGAAGACAGCTTCTTGCAGCGATACAGCTCAGACCCCACAGGCGCCTTGACTG 3235
P--I--K--E--D--S--F--L--Q--R--Y--S--S--D--P--T--G--A--L--T-- 1078

AGGACAGCATAGACGACACCTTCCTCCCAGTGCCTGAATACATAAACCAGTCCGTTCCCA 3295
E--D--S--I--D--D--T--F--L--P--V--P--E--Y--I--N--Q--S--V--P-- 1098

AAAGGCCCGCTGGCTCTGTGCAGAATCCTGTCTATCACAATCAGCCTCTGAACCCCGCGC 3355
K--R--P--A--G--S--V--Q--N--P--V--Y--H--N--Q--P--L--N--P--A-- 1118

CCAGCAGAGACCCACACTACCAGGACCCCCACAGCACTGCAGTGGGCAACCCCGAGTATC 3415
P--S--R--D--P--H--Y--Q--D--P--H--S--T--A--V--G--N--P--E--Y-- 1138

TCAACACTGTCCAGCCCACCTGTGTCAACAGCACATTCGACAGCCCTGCCCACTGGGCCC 3475
L--N--T--V--Q--P--T--C--V--N--S--T--F--D--S--P--A--H--W--A-- 1158

AGAAAGGCAGCCACCAAATTAGCCTGGACAACCCTGACTACCAGCAGGACTTCTTTCCCA 3535
Q--K--G--S--H--Q--I--S--L--D--N--P--D--Y--Q--Q--D--F--F--P-- 1178

AGGAAGCCAAGCCAAATGGCATCTTTAAGGGCTCCACAGCTGAAAATGCAGAATACCTAA 3595
K--E--A--K--P--N--G--I--F--K--G--S--T--A--E--N--A--E--Y--L-- 1198

GGGTCGCGCCACAAAGCAGTGAATTTATTGGAGCATGA 3633 (SEQ ID NO 511)
R--V--A--P--Q--S--S--E--F--I--G--A--*- 1210 (SEQ ID NO 512)
```

FIG. 6

A

```
                                                         SEQ ID NO
                     activation loop
L858R   KTPQHVKITDFGRAKLLGAEEKEYH    870       477
EGFR    KTPQHVKITDFGLAKLLGAEEKEYH    870       478
BRAF    HEDLTVKIGDFGLATVKSRWSGSHQ    608       479
             *** *****
```

B

```
                        P-loop
G719S   ETEFKKIKVLSSGAFGTVYKGLWIP    733       480
EGFR    ETEFKKIKVLGSGAFGTVYKGLWIP    733       481
BRAF    DGQITVGQRIGSGSFGTVYKGKWHG    477       482
                  *** ******* *
```

C

```
        742     |750|752
Del-1   VAIK-----T-SPKANKEILDEAYV    765       483
Del-2   VAIKELREAT---------LDEAYV    765       484
Del-3   VAIKE---PT-SPKANKEILDEAYV    765       485
Del-4   VAIKE-------SKANKEILDEAYV    765       486
Del-5   VAIKV-------PKANKEILDEAYV    765       487
EGFR    VAIKELREAT-SPKANKEILDEAYV    765       488
BRAF    VAVKMLNVTAPTPQQLQAFKNEVGV    503       489
        ** *                 *  *
```

FIG. 7

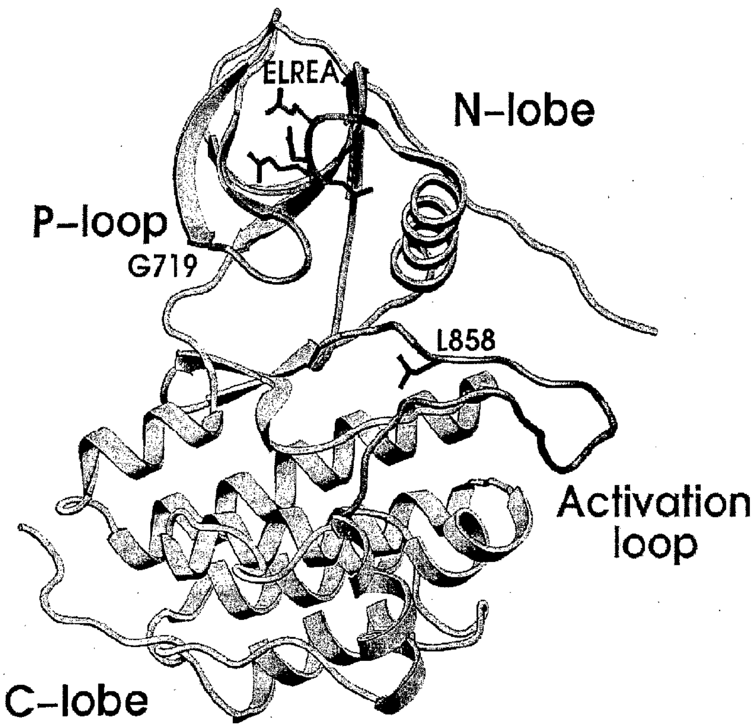


FIG. 8

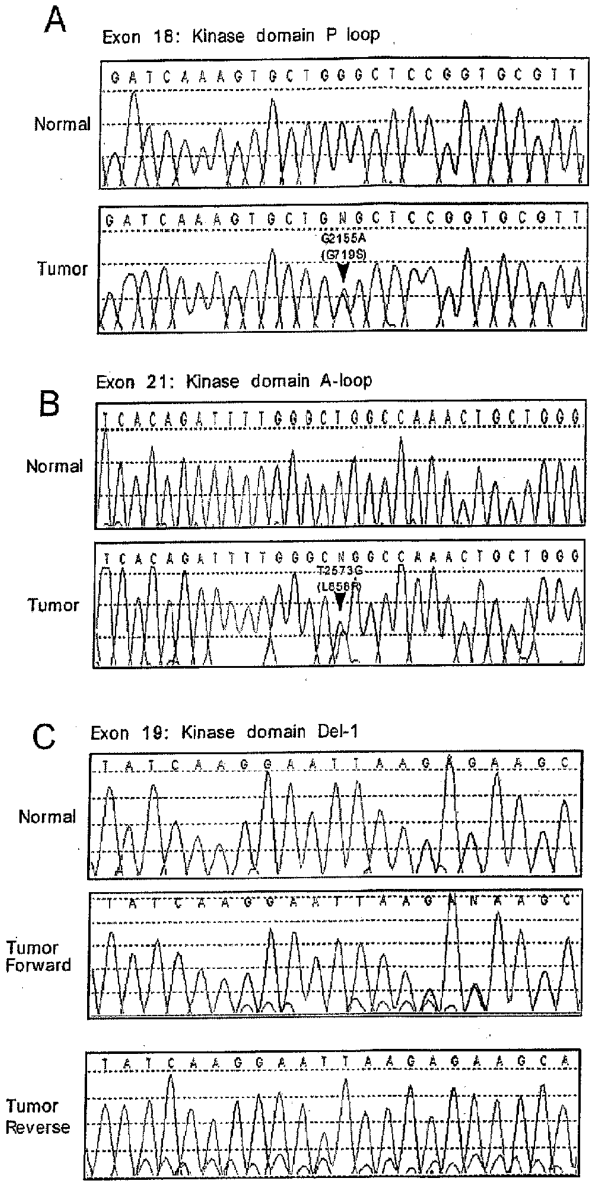
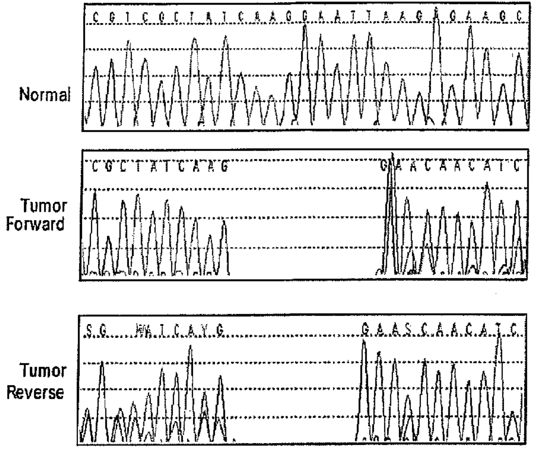

A
Exon 18: Kinase domain P loop
Normal
GATCAAAGTGCTGGGCTCCGGTGCGTT
Tumor
GATCAAAGTGCTGNGCTCCGGTGCGTT
G2155A
(G719S)
B
Exon 21: Kinase domain A-loop
Normal
TCACAGATTTTGGGCTGGCCAAACTGCTGGG
Tumor
TCACAGATTTTGGGCNGGCCAAACTGCTGGG
T2573G
(L858R)
C
Exon 19: Kinase domain Del-1
Normal
TATCAAGGAATTAAGAGAAGC
Tumor
Forward
TATCAAGGAATTAAGANAAGC
Tumor
Reverse
TATCAAGGAATTAAGAGAAGCA

FIG. 8 (cont.)

D Exon 19: Kinase domain Del-3



E Exon 19: Kinase domain Del-4

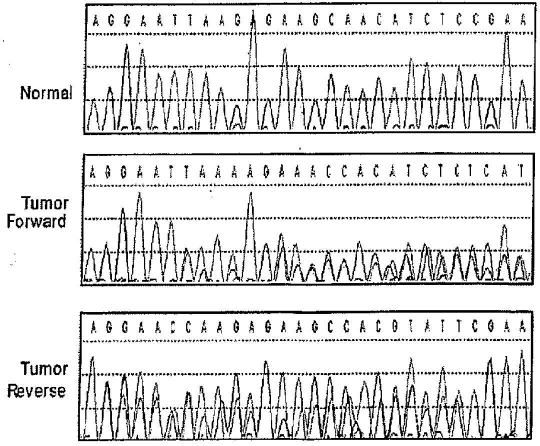


FIG. 8 (cont.)

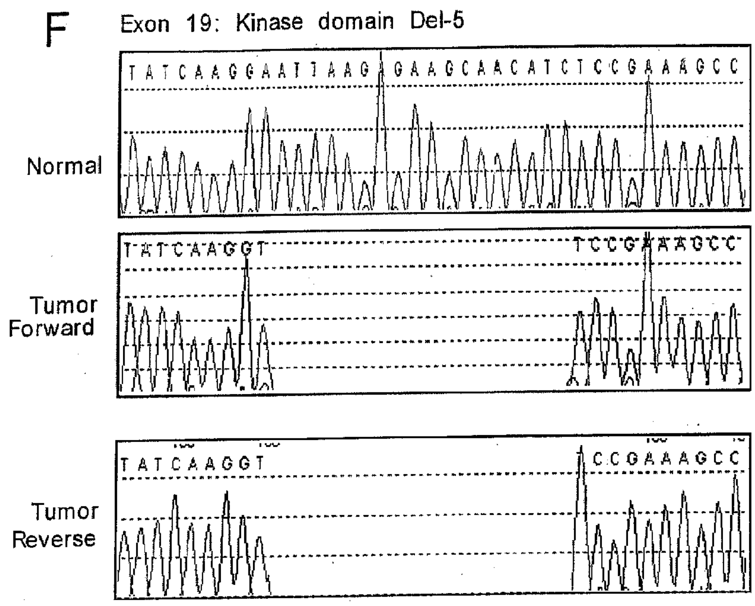
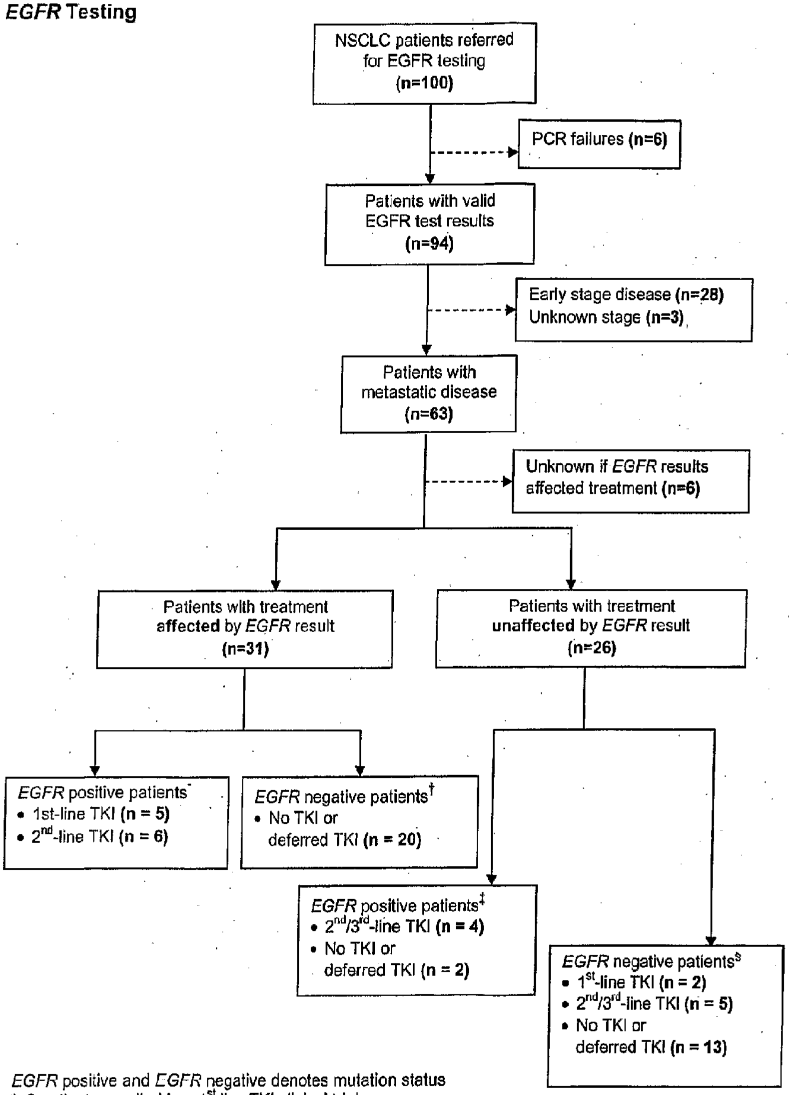

F
Exon 19: Kinase domain Del-5
TATCAAGGAATTAAGAGAAGCAACATCTCCGAAAGCC
Normal
TATCAAGGT
TCCGAAAGCC
Tumor Forward
TATCAAGGT
CCGAAAGCC
Tumor Reverse

FIG 9

Alignment of Abl1 to EGFR

```
          712                                                                         790
Abl1(242) ITMKHKLGGGQYGEVYEGVWK----KYSLTVAVKTLKEDT--MEVEEFLKEAAVMKEIKHPNLVQLLGVCTREPPFYIIT
EGFR(712) FKKIKVLGSGAFGTVYKGLWIPEGEKVKIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGICLTS-TVQLIT
Consensus       LG G FG VY GLW     K  I VAIK LKE T      E L EA VM  I  P L  LLGIC       IIT
          791                                                                         869
Abl1(316) EFMTYGNLLDYLRECNRQEVNAVVLLYMATQISSAMEYLEKKNFIHRDLAARNCLVGENHLVKVADFGLSRLMTGDTYTA
EGFR(791) QLMPFGCLLDYVREHKDN-IGSQYLLNWCVQIAKGMNYLEDRRLVHRDLAARNVLVKTPQHVKITDFGLAKLLGAEEKEY
Consensus   M FG LLDYLRE   N I A  LL    QIA AM YLE K  IHRDLAARN LV     VKI DFGLAKLL AD
          870                                                                         949
Abl1(396) HA-GAKFPIKWTAPESLAYNKFSIKSDVWAFGVLLWEIATYGMSPYPGIDLSQVYELLEKDYRMERPEGCPEKVYELMRA
EGFR(870) HAEGGKVPIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSILEKGERLPQPPICTIDVYMIMVK
Consensus HA GAK PIKW A ESI H  FS  SDVWAFGV LWEI TFG  PY GI  S I  ILEK  RL  P  C   VY IM
          950                           979
Abl1(475) CWQWNPSDRPSFAEIHQAF----------- (SEQ ID NO 491)
EGFR(950) CWMIDADSRPKFRELIIEFSKMARDPQRYL  (SEQ ID NO 492)
Consensus CW      RP F EI   F
```

Figure 10 The Decision Making Process for Patients with Metastatic NSCLC Undergoing *EGFR* Testing



EGFR positive and *EGFR* negative denotes mutation status
* 2 patients enrolled in a 1st line TKI clinical trial
† 2 patients receiving TKI prior to testing
‡ 3 patients receiving TKI prior to testing
§ 1 patient enrolled in 1st line TKI clinical trial and 4 patients receiving TKI prior to testing

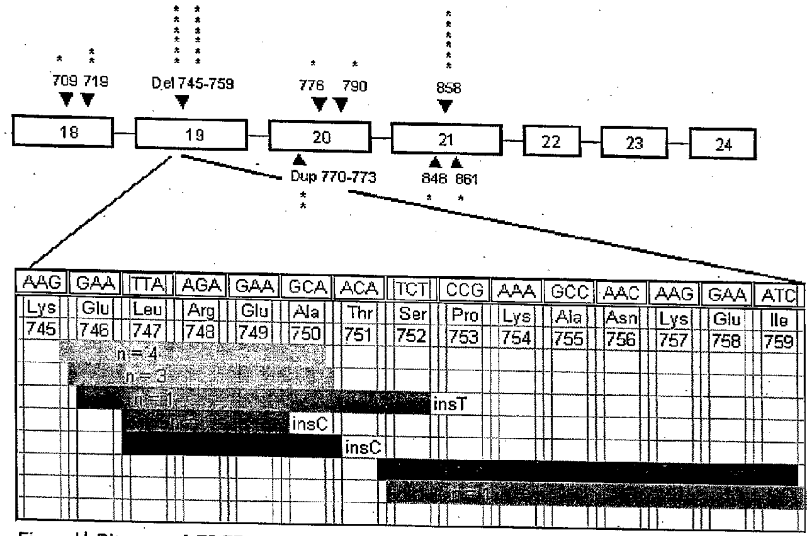


Figure 11 Diagram of *EGFR* exons 18-24, which encode the TK region (not to scale). Arrows depict the location of identified mutations. Astericks denote the number of patients with mutations at each location. Blow-up diagram depicts the overlap of the exon 19 deletions, and the number of patients (n) with each deletion.