



(51) International Patent Classification:
G01N 33/574 (2006.01) *G01N 33/52* (2006.01)
C12Q 1/68 (2006.01)

VALL D'HEBRON UNIVERSITY HOSPITAL
[ES/ES]; Passeig De La D' Hebron, 119-129, E-08035
Barcelona (ES).

(21) International Application Number:
PCT/US2011/060385

(72) Inventors; and

(22) International Filing Date:
11 November 2011 (11.11.2011)

(75) Inventors/Applicants (*for US only*): **MEYERSON, Matthew** [US/US]; 307 Independence Road, Concord, MA 01742 (US). **BASS, Adam** [US/US]; 46 Fayette Street, Cambridge, MA 02139 (US). **LAWRENCE, Michael** [US/US]; 11 Norris Street, Cambridge, MA 02140 (US). **TABERNERO, Josep** [ES/ES]; 307 Independence Road, Barcelona (ES). **BASLGA, Jose** [ES/ES]; 11 Norris Street, Barcelona (ES). **CHO, Jeonghee** [KR/KR]; Yimgwang Apt 311-801, Seohyeondong, Bundang-gu, Seongnam-si, Gyeonggi-do 463-766 (KR).

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/413,158 12 November 2010 (12.11.2010) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 61/413,158 (CIP)
Filed on 12 November 2010 (12.11.2010)

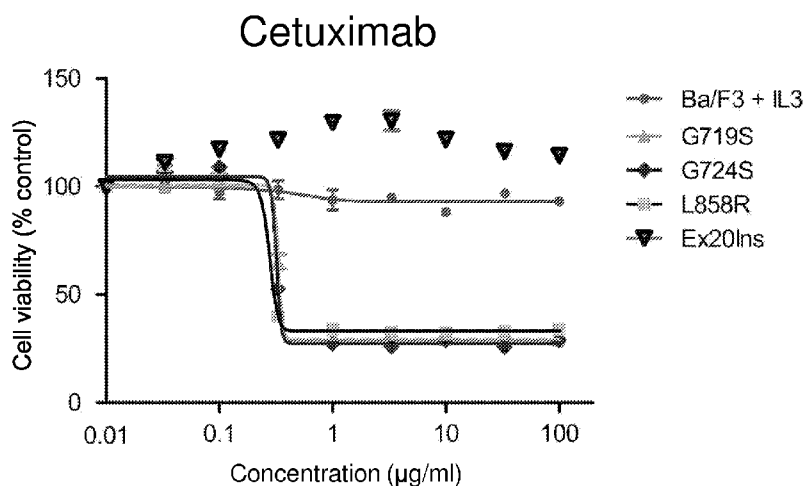
(74) Agents: **ELRIFI, Ivor, R.** et al.; Mintz Levin Cohn Ferris Glovsky And Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

(71) Applicants (*for all designated States except US*): **THE BROAD INSTITUTE OF MIT AND HARVARD** [US/US]; The Broad Institute Of Mit And Harvard, 7 Cambridge Center #7034, Cambridge, MA 02142 (US). **DANA-FARBER CANCER INSTITUTE, INC.** [US/US]; 450 Brookline Avenue, Boston, MA 02215 (US).

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

[Continued on next page]

(54) Title: METHODS OF PREDICTING RESPONSE TO EGFR ANTIBODY THERAPY



(57) Abstract: This present invention provides methods of treating of accessing/monitoring the responsiveness of a cancer cell to EGFR antibody therapy.

FIG. 3



HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ,

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

METHODS OF PREDICTING RESPONSE TO EGFR ANTIBODY THERAPY

RELATED REFERENCE

This application claims priority to, and the benefit of, U.S.S.N. 61/413,158, filed
5 November 12, 2010, which is incorporated herein by reference in its entirety.

FIELD OF THE INVENTION

The invention relates to generally to the predicting a subjects response to EGFR
10 antibody therapy.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

This invention was made with government support under R01CA116020 awarded by
the National Institutes of Health and under U54HG003067 awarded by the National Human
Genome Research Institute. The government has certain rights in the invention.

BACKGROUND OF THE INVENTION

Colorectal adenocarcinoma is it the second leading cause of cancer death in the
United States, leading to approximately 55,000 deaths per year. Although there have been
market improvements in our ability to screen for this disease enabling diagnosis when the
20 cancer is curable by surgery, there are still many people who eventually succumb to
metastatic disease every fear. Fortunately, improvements in cytotoxic chemotherapy have led
to survival improvements for these patients. A useful adjunct to cytotoxic therapy has been
the introduction of targeted biologic agents for the treatment of metastatic colorectal cancer.
One such agent is Cetuximab, a monoclonal antibody targeting the Epidermal Growth Factor
25 Receptor (EGFR). The combination of Cetuximab and irinotecan was shown to increase the
likelihood of clinical response and lead to a significant survival benefit when compared to
chemotherapy alone in patients who had already failed initial cytotoxic chemotherapy. In the
pivotal trial leading to approval of this agent, treatment with Cetuximab alone led to a 10%
response rate whereas combination of the antibody with chemotherapy improved response
30 rate to approximately 20%. Based on these results, use of Cetuximab is now FDA approved
in combination with chemotherapy in the second line setting after failure of initial therapy.
Although Cetuximab is now widely used in patients with metastatic colorectal cancer, there is

a great need to develop a predictive test to enable the drug to be given to those patients with the greatest likelihood of response as early as possible.

SUMMARY OF THE INVENTION

5 The invention features method of accessing the effectiveness of an EGFR antibody therapy treatment regimen of a subject having cancer by obtaining a sample from the subject and detecting the presence or absence of a G724S mutation, a G719S mutation, or a L858R mutation in the EGFR polypeptide. The presence of the mutation indicates the subject is responsive to EGFR antibody therapy treatment. EGFR antibodies suitable for therapy
10 include for example Cetuximab, Panitumumab, Zalutumumab, Nimotuzumab, Necitumumab and Matuzumab. The cancer is any cancer capable of being treated with an EGFR antibody. For example, the cancer is colorectal adenocarcinoma or head and neck cancer. In some aspects the subject has not received treatment for the cancer. Alternatively, the subject has received treatment for the cancer.

15 The method is implemented in a flow-cytometry (FC), immuno-histochemistry (IHC), immuno-fluorescence (IF) assay format or a polymerase chain reaction (PCR) sequencing assay format.

 Also provided by the invention is a kit for the detection of mutant (G724S) EGFR polypeptide in a biological sample, the kit containing a reagent specifically binds to or detects
20 a mutant EGFR polypeptide but does not bind to or detect wild type EGFR and one or more secondary reagents.

 Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can
25 be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and are not intended to be limiting.

30 Other features and advantages of the invention will be apparent from the following detailed description and claims.

BRIEF DESCRIPTION OF DRAWINGS

Figure 1 depicts graphs establishing that colon cancer derived G719S and G724S EGFR mutants are oncogenic and dependent on dimerization for transforming activity. Panels A- D shows growth of colonies of wild type (A), L858R mutant (B), G724S mutant (C) and G719S mutant (D) in soft agar. Panel E shows the relative colony numbers of G724S dimmers, G724S/L704N dimmers and G724S/I941R dimmers. Panel F shows the relative colony numbers of G719S dimmers, G719S/L704N dimmers and G719S/I941R dimmers.

Figure 2 depicts a graph demonstrating that G719S and G724S EGFR mutants are not as sensitive as L858R mutant to Erlotinib.

Figure 3 depicts a graph demonstrating that G719S and G724S EGFR mutants are as sensitive as L858R mutant to Cetuximab.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to the identification of biomarkers associated with the responsiveness to EGFR antibody therapy. Specifically, the biomarkers are somatic mutations in the EGFR gene leading to an amino acid substitution at position 724, 719 or 858. The amino acid at position 724 is in the nucleotide binding loop of the EGFR polypeptide. In particular, the mutations result in a glycine to a serine at position 724 (G724S) and at position 719 (G719S) and a leucine to an arginine at position 858 (L858R).

Anti-EGFR antibodies suitable for human administration are known in the art and include for example, Cetuximab, Panitumumab, Zalutumumab, Nimotuzumab, Necitumumab and Matuzumab.

Cetuximab (also known as Erbitux[®]) is a chimeric (mouse/human) monoclonal antibody that is an epidermal growth factor receptor (EGFR) inhibitor useful for treatment of colorectal cancer and head and neck cancer.

Cetuximab is indicated for the treatment of patients with EGFR expressing, KRAS wild-type metastatic colorectal cancer in combination with chemotherapy or as a single agent in patients who have failed in oxaliplatin- or irinotecan- based therapy and who are intolerant to irinotecan. Cetuximab is also indicated for use in combination with radiation therapy for treating squamous cell carcinoma of the head and neck (SCCHN) or as a single agent in patients who have had prior platinum-based therapy.

Panitumumab (Vectibix) is indicated for the treatment of EGFR-expressing metastatic colorectal cancer with disease progression despite prior treatment.

Zalutumumab (HuMax-EGFr) is indicated for the treatment of patients suffering from SCCHN who have failed standard therapies and have no other options. The drug has undergone pre-clinical and Phase I and II studies and is also in Phases I and II for SCCHN front-line with chemo-radiation and SCCHN with radiation. Additionally, a Phase II is under way for SCCHN and Phase III studies are also being performed for SCCHN and SCCHN front-line with radio therapy. A phase III study reported a non significant improvement in overall survival, and a significant improvement in Progression-free survival.

Nimotuzumab (BIOMAb EGFR) is a chimeric monoclonal antibody used to treat squamous cell carcinomas of the head and neck, recurrent/refractory high grade malignant astrocytoma and glioma.

While, EGFR antibody therapy alone or in combination has shown a clinical benefit in those unresponsive to other treatment, however the response rate is low. Accordingly a need exist is identify biomarkers that are predictive of responsiveness to EGFR antibody therapy therapy.

Accordingly, the invention provides methods of determining the responsiveness, *e.g.*, sensitivity or resistance, of a cancer cell to EGFR antibody therapy. These methods are also useful for monitoring subjects undergoing treatments and therapies for cancer such as for example colorectal adenocarcinoma or head and neck cancer, and for selecting therapies and treatments that would be efficacious in subjects having cancer, wherein selection and use of such treatments and therapies slow the progression of cancer. More specifically, the invention provides methods of determining the whether a patient with colorectal adenocarcinoma will be responsive to EGFR antibody therapy.

Because of the remarkable progress in genomic technology now allows us much greater power for deep discovery of genomic alterations in cancer that we could use these technologies to identify positive predictive markers of EGFR antibody therapy response that had been previously missed. To augment our power for discovery, our initial starting material for genomic discovery was a patient sample with the highest possible likelihood of identification of any genomic projector response. The patients whose tumor we chose to profile for this study was one whose tumor was obtained prior to enrollment in a specific clinical trial whereby patients with metastatic colorectal adenocarcinoma were treated with

the use of single agent Cetuximab prior to the receipt of other cytotoxic chemotherapy. In this study, patients received six weeks of Cetuximab and then were re-staged to gauge the response to Cetuximab alone prior to the receipt of Cetuximab in combination with standard chemotherapy. Using tumor DNA from a patient who developed a clinical partial response
5 from Cetuximab treatment alone, we performed deep sequencing to identify a genomic alteration leading to dependency upon EGFR activity.

Remarkably, we identified three somatic mutations, G724S, G719S and L858R, one of which is in the P-loop of the tyrosine kinase of EGFR in this patient. It is of course notable that there has been such a substantial prior effort to identify somatic coding mutations
10 of EGFR in colorectal cancer patients, an effort which has demonstrated that these mutations are quite uncommon. Significantly though, these prior studies used a traditional sequencing a method is much less sensitive than the next-generation sequencing that we performed in our study. In fact, the patient whose tumor was sequenced demonstrated mutant EGFR reads in only 10 of the 42 reads covering the mutation. Given this relatively low fraction of alleles in
15 the DNA sample likely harboring the activating mutation, it could easily be imagined that similar events in other patients could have been missed by traditional sequencing technologies due to the relatively low number of tumor alleles and DNA mixture. Should this be the case, further sequencing of additional clinical responders using these more sensitive mutation-sequencing methods may identify EGFR events that may be a similarly likely basis
20 for response to therapy.

The new data presented herein is indicative that the prior understanding of the rationale for response to Cetuximab and EGFR antibody therapy in general in patients with colorectal adenocarcinoma may understate the degree of genomic aberrations of EGFR in this disease. Thus, we believe that continued deep sequencing in this disease will yield similar
25 events and that in the future application of next generation sequencing to patients with colorectal adenocarcinoma may discovery further mutations in EGFR for the identification of those patients who should be treated with initial EGFR antibody therapy therapy, either alone or in combination with cytotoxic therapy. Thus, this finding has dramatic potential implications for the therapy of patients with colorectal carcinoma. It may serve as both a
30 positive and negative predictive test and thus allow clinicians to better focus the use of this expensive and toxic agent to that subset of the population the greatest potential chance of benefit as early as possible.

Definitions

“Accuracy” refers to the degree of conformity of a measured or calculated quantity (a test reported value) to its actual (or true) value. Clinical accuracy relates to the proportion of true outcomes (true positives (TP) or true negatives (TN) versus misclassified outcomes (false positives (FP) or false negatives (FN)), and may be stated as a sensitivity, specificity, positive predictive values (PPV) or negative predictive values (NPV), or as a likelihood, odds ratio, among other measures.

“Biomarker” in the context of the present invention encompasses, without limitation, proteins, nucleic acids, and metabolites, together with their polymorphisms, mutations, variants, modifications, subunits, fragments, protein-ligand complexes, and degradation products, protein-ligand complexes, elements, related metabolites, and other analytes or sample-derived measures. Biomarkers can also include mutated proteins or mutated nucleic acids. Biomarkers also encompass non-blood borne factors or non-analyte physiological markers of health status, such as “clinical parameters” defined herein, as well as “traditional laboratory risk factors”, also defined herein. Biomarkers also include any calculated indices created mathematically or combinations of any one or more of the foregoing measurements, including temporal trends and differences. Where available, and unless otherwise described herein, biomarkers which are gene products are identified based on the official letter abbreviation or gene symbol assigned by the international Human Genome Organization Naming Committee (HGNC) and listed at the date of this filing at the US National Center for Biotechnology Information (NCBI) web site.

A “Clinical indicator” is any physiological datum used alone or in conjunction with other data in evaluating the physiological condition of a collection of cells or of an organism. This term includes pre-clinical indicators.

“Clinical parameters” encompasses all non-sample or non-analyte biomarkers of subject health status or other characteristics, such as, without limitation, age (Age), ethnicity (RACE), gender (Sex), or family history (FamHX).

“FN” is false negative, which for a disease state test means classifying a disease subject incorrectly as non-disease or normal.

“FP” is false positive, which for a disease state test means classifying a normal subject incorrectly as having disease.

A “formula,” “algorithm,” or “model” is any mathematical equation, algorithmic, analytical or programmed process, or statistical technique that takes one or more continuous or categorical inputs (herein called “parameters”) and calculates an output value, sometimes referred to as an “index” or “index value.” Non-limiting examples of “formulas” include sums, ratios, and regression operators, such as coefficients or exponents, biomarker value transformations and normalizations (including, without limitation, those normalization schemes based on clinical parameters, such as gender, age, or ethnicity), rules and guidelines, statistical classification models, and neural networks trained on historical populations. Of particular use in combining biomarkers are linear and non-linear equations and statistical classification analyses to determine the relationship between biomarkers detected in a subject sample and the subject’s responsiveness to chemotherapy. In panel and combination construction, of particular interest are structural and synactic statistical classification algorithms, and methods of risk index construction, utilizing pattern recognition features, including established techniques such as cross-correlation, Principal Components Analysis (PCA), factor rotation, Logistic Regression (LogReg), Linear Discriminant Analysis (LDA), Eigengene Linear Discriminant Analysis (ELDA), Support Vector Machines (SVM), Random Forest (RF), Recursive Partitioning Tree (RPART), as well as other related decision tree classification techniques, Shrunk Centroids (SC), StepAIC, Kth-Nearest Neighbor, Boosting, Decision Trees, Neural Networks, Bayesian Networks, Support Vector Machines, and Hidden Markov Models, among others. Other techniques may be used in survival and time to event hazard analysis, including Cox, Weibull, Kaplan-Meier and Greenwood models well known to those of skill in the art. Many of these techniques are useful as forward selection, backwards selection, or stepwise selection, complete enumeration of all potential panels of a given size, genetic algorithms, or they may themselves include biomarker selection methodologies in their own technique. These may be coupled with information criteria, such as Akaike’s Information Criterion (AIC) or Bayes Information Criterion (BIC), in order to quantify the tradeoff between additional biomarkers and model improvement, and to aid in minimizing overfit. The resulting predictive models may be validated in other studies, or cross-validated in the study they were originally trained in, using such techniques as Bootstrap, Leave-One-Out (LOO) and 10-Fold cross-validation (10-Fold CV). At various steps, false discovery rates may be estimated by value permutation according to techniques known in the art. A “health economic utility function” is a formula that is derived from a

combination of the expected probability of a range of clinical outcomes in an idealized applicable patient population, both before and after the introduction of a diagnostic or therapeutic intervention into the standard of care. It encompasses estimates of the accuracy, effectiveness and performance characteristics of such intervention, and a cost and/or value measurement (a utility) associated with each outcome, which may be derived from actual health system costs of care (services, supplies, devices and drugs, etc.) and/or as an estimated acceptable value per quality adjusted life year (QALY) resulting in each outcome. The sum, across all predicted outcomes, of the product of the predicted population size for an outcome multiplied by the respective outcome's expected utility is the total health economic utility of a given standard of care. The difference between (i) the total health economic utility calculated for the standard of care with the intervention versus (ii) the total health economic utility for the standard of care without the intervention results in an overall measure of the health economic cost or value of the intervention. This may itself be divided amongst the entire patient group being analyzed (or solely amongst the intervention group) to arrive at a cost per unit intervention, and to guide such decisions as market positioning, pricing, and assumptions of health system acceptance. Such health economic utility functions are commonly used to compare the cost-effectiveness of the intervention, but may also be transformed to estimate the acceptable value per QALY the health care system is willing to pay, or the acceptable cost-effective clinical performance characteristics required of a new intervention.

For diagnostic (or prognostic) interventions of the invention, as each outcome (which in a disease classifying diagnostic test may be a TP, FP, TN, or FN) bears a different cost, a health economic utility function may preferentially favor sensitivity over specificity, or PPV over NPV based on the clinical situation and individual outcome costs and value, and thus provides another measure of health economic performance and value which may be different from more direct clinical or analytical performance measures. These different measurements and relative trade-offs generally will converge only in the case of a perfect test, with zero error rate (a.k.a., zero predicted subject outcome misclassifications or FP and FN), which all performance measures will favor over imperfection, but to differing degrees.

“Measuring” or “measurement,” or alternatively “detecting” or “detection,” means assessing the presence, absence, quantity or amount (which can be an effective amount) of either a given substance within a clinical or subject-derived sample, including the derivation

of qualitative or quantitative concentration levels of such substances, or otherwise evaluating the values or categorization of a subject's non-analyte clinical parameters.

“Negative predictive value” or “NPV” is calculated by $TN/(TN + FN)$ or the true negative fraction of all negative test results. It also is inherently impacted by the prevalence of the disease and pre-test probability of the population intended to be tested.

See, *e.g.*, O'Marcaigh AS, Jacobson RM, “Estimating The Predictive Value Of A Diagnostic Test, How To Prevent Misleading Or Confusing Results,” Clin. Ped. 1993, 32(8): 485-491, which discusses specificity, sensitivity, and positive and negative predictive values of a test, *e.g.*, a clinical diagnostic test. Often, for binary disease state classification approaches using a continuous diagnostic test measurement, the sensitivity and specificity is summarized by Receiver Operating Characteristics (ROC) curves according to Pepe et al, “Limitations of the Odds Ratio in Gauging the Performance of a Diagnostic, Prognostic, or Screening Marker,” Am. J. Epidemiol 2004, 159 (9): 882-890, and summarized by the Area Under the Curve (AUC) or c-statistic, an indicator that allows representation of the sensitivity and specificity of a test, assay, or method over the entire range of test (or assay) cut points with just a single value. See also, *e.g.*, Shultz, “Clinical Interpretation Of Laboratory Procedures,” chapter 14 in Teitz, Fundamentals of Clinical Chemistry, Burtis and Ashwood (eds.), 4th edition 1996, W.B. Saunders Company, pages 192-199; and Zweig et al., “ROC Curve Analysis: An Example Showing The Relationships Among Serum Lipid And Apolipoprotein Concentrations In Identifying Subjects With Coronary Artery Disease,” Clin. Chem., 1992, 38(8): 1425-1428. An alternative approach using likelihood functions, odds ratios, information theory, predictive values, calibration (including goodness-of-fit), and reclassification measurements is summarized according to Cook, “Use and Misuse of the Receiver Operating Characteristic Curve in Risk Prediction,” Circulation 2007, 115: 928-935. Finally, hazard ratios and absolute and relative risk ratios within subject cohorts defined by a test are a further measurement of clinical accuracy and utility. Multiple methods are frequently used to defining abnormal or disease values, including reference limits, discrimination limits, and risk thresholds.

“Analytical accuracy” refers to the reproducibility and predictability of the measurement process itself, and may be summarized in such measurements as coefficients of variation, and tests of concordance and calibration of the same samples or controls with

different times, users, equipment and/or reagents. These and other considerations in evaluating new biomarkers are also summarized in Vasan, 2006.

“Performance” is a term that relates to the overall usefulness and quality of a diagnostic or prognostic test, including, among others, clinical and analytical accuracy, other analytical and process characteristics, such as use characteristics (*e.g.*, stability, ease of use), health economic value, and relative costs of components of the test. Any of these factors may be the source of superior performance and thus usefulness of the test, and may be measured by appropriate “performance metrics,” such as AUC, time to result, shelf life, etc. as relevant.

“Positive predictive value” or “PPV” is calculated by $TP/(TP+FP)$ or the true positive fraction of all positive test results. It is inherently impacted by the prevalence of the disease and pre-test probability of the population intended to be tested.

“Risk” in the context of the present invention, relates to the probability that an event will occur over a specific time period, as in the responsiveness to treatment, cancer recurrence or survival and can mean a subject’s “absolute” risk or “relative” risk. Absolute risk can be measured with reference to either actual observation post-measurement for the relevant time cohort, or with reference to index values developed from statistically valid historical cohorts that have been followed for the relevant time period. Relative risk refers to the ratio of absolute risks of a subject compared either to the absolute risks of low risk cohorts or an average population risk, which can vary by how clinical risk factors are assessed. Odds ratios, the proportion of positive events to negative events for a given test result, are also commonly used (odds are according to the formula $p/(1-p)$ where p is the probability of event and $(1-p)$ is the probability of no event) to no-conversion.

“Risk evaluation” or “evaluation of risk” in the context of the present invention encompasses making a prediction of the probability, odds, or likelihood that an event or disease state may occur, the rate of occurrence of the event or conversion from one disease state. Risk evaluation can also comprise prediction of future clinical parameters, traditional laboratory risk factor values, or other indices of cancer, either in absolute or relative terms in reference to a previously measured population. The methods of the present invention may be used to make continuous or categorical measurements of the responsiveness to treatment thus diagnosing and defining the risk spectrum of a category of subjects defined as being responders or non-responders. In the categorical scenario, the invention can be used to discriminate between normal and other subject cohorts at higher risk for responding. Such

differing use may require different HNCMARKER combinations and individualized panels, mathematical algorithms, and/or cut-off points, but be subject to the same aforementioned measurements of accuracy and performance for the respective intended use.

A “sample” in the context of the present invention is a biological sample isolated from a subject and can include, by way of example and not limitation, tissue biopies, whole blood, serum, plasma, blood cells, endothelial cells, lymphatic fluid, ascites fluid, interstitial fluid (also known as “extracellular fluid” and encompasses the fluid found in spaces between cells, including, *inter alia*, gingival crevicular fluid), bone marrow, cerebrospinal fluid (CSF), saliva, mucous, sputum, sweat, urine, or any other secretion, excretion, or other bodily fluids.

A “sample” may include a single cell or multiple cells or fragments of cells. The sample is also a tissue sample. The sample is or contains a circulating endothelial cell or a circulating tumor cell. The sample includes a primary tumor cell, primary tumor, a recurrent tumor cell, or a metastatic tumor cell.

“Sensitivity” is calculated by $TP/(TP+FN)$ or the true positive fraction of disease subjects.

“Specificity” is calculated by $TN/(TN+FP)$ or the true negative fraction of non-disease or normal subjects.

[0001] By “statistically significant”, it is meant that the alteration is greater than what might be expected to happen by chance alone (which could be a “false positive”). Statistical significance can be determined by any method known in the art. Commonly used measures of significance include the p-value, which presents the probability of obtaining a result at least as extreme as a given data point, assuming the data point was the result of chance alone. A result is considered highly significant at a p-value of 0.05 or less. Preferably, the p-value is 0.04, 0.03, 0.02, 0.01, 0.005, 0.001 or less.

A “subject” in the context of the present invention is preferably a mammal. The mammal can be a human, non-human primate, mouse, rat, dog, cat, horse, or cow, but are not limited to these examples. Mammals other than humans can be advantageously used as subjects that represent animal models of cancer. A subject can be male or female.

“TN” is true negative, which for a disease state test means classifying a non-disease or normal subject correctly.

“TP” is true positive, which for a disease state test means correctly classifying a disease subject.

“Traditional laboratory risk factors” correspond to biomarkers isolated or derived from subject samples and which are currently evaluated in the clinical laboratory and used in traditional global risk assessment algorithms. Traditional laboratory risk factors for tumor recurrence include for example Proliferative index, tumor infiltrating lymphocytes. Other
5 traditional laboratory risk factors for tumor recurrence known to those skilled in the art.

Methods And Uses Of The Invention

The methods disclosed herein are used with subjects undergoing treatment and/or therapies for cancer, subjects who are at risk for developing a reoccurrence of cancer (*e.g.* colorectal adenocarcinoma or head and neck cancer), and subjects who have been diagnosed
10 with cancer. The methods of the present invention are to be used to monitor or select a treatment regimen for a subject who has cancer, and to evaluate the predicted survivability and/or survival time of a cancer-diagnosed subject. Treatment regimens include EGFR antibody therapies such as Cetuximab, Panitumumab, Zalutumumab, Nimotuzumab,
15 Necitumumab and Matuzumab.

Responsiveness (*e.g.*, resistance or sensitivity) of a cell to EGFR antibody therapy is determined by detecting a mutation associated with responsiveness to EGFR antibody therapy in a test sample (*e.g.*, a subject derived sample). The mutation associated with responsiveness to EGFR antibody therapy includes a somatic mutation in the EGFR
20 gene leading to an amino acid substitution at position 724, at position 719 or at position 858 of the EGFR polypeptide. Specifically the mutation results in a glycine to serine at positions 724 (G724S) and 719 (G719S) and a leucine to an arginine at position 858 (L858R). The mutation associated with responsiveness to EGFR antibody therapy is referred herein as the EGFR G724S mutation, the EGFR G719S mutation or the EGFR L858R mutation. The
25 presence of the EGFR G724S mutation, the EGFR G719S mutation, the EGFR L858R mutation, or any combination thereof indicates the cell will be responsive (*i.e.*, sensitive) to EGFR antibody therapy. In contrast, the absence of these mutations indicates the cell will be non- responsive (*i.e.*, resistant) to EGFR antibody therapy. For example, cells expressing G724S mutant or G719S mutant are as sensitive as L858R mutant to Cetuximab (Figure 3),
30 but less sensitive to Erlotinib (Figure 2).

The cell is for example a cancer cell. Optionally, the cancer is a colorectal adenocarcinoma cancer or a head or neck cancer such as cancer of the nasal cavity, sinuses, lips, mouth, salivary glands, throat, or larynx

By resistance it is meant that a cell fails to respond to an agent. For example,
5 resistance to EGFR antibody therapy means the cell is not damaged or killed by the drug. By sensitivity it is meant that that the cell responds to an agent. For example, sensitivity to EGFR antibody therapy means the cell is damaged or killed by the drug.

The methods of the present invention are useful to treat, alleviate the symptoms of, monitor the progression of or delay the onset of cancer.

10 Preferably, the methods of the present invention are used to identify and/or diagnose subjects who are asymptomatic for a cancer recurrence. "Asymptomatic" means not exhibiting the traditional symptoms.

The methods of the present invention are also useful to identify and/or diagnose subjects already at higher risk of developing a head and neck cancer or based on solely on the
15 traditional risk factors.

Identification of the EGFR G724S G719S or L858R mutation allows for the determination of whether a subject will derive a benefit from a particular course of treatment. In this method, a biological sample is provided from a subject before undergoing treatment, *e.g.*, EGFR antibody therapy or combinations thereof. By "derive a benefit" it is meant that
20 the subject will respond to the course of treatment. By responding it is meant that the treatment decreases in size, prevalence, or metastatic potential of a cancer (*e.g.* colorectal adenocarcinoma or head and neck cancer) in a subject. When treatment is applied prophylactically, "responding" means that the treatment retards or prevents a cancer (*e.g.* colorectal adenocarcinoma or head and neck cancer) recurrence from forming or retards,
25 prevents, or alleviates a symptom. Assessment of cancers are made using standard clinical protocols.

The present invention can also be used to screen patient or subject populations in any number of settings. For example, a health maintenance organization, public health entity or school health program can screen a group of subjects to identify those requiring interventions,
30 as described above, or for the collection of epidemiological data. Insurance companies (*e.g.*, health, life or disability) may screen applicants in the process of determining coverage or pricing, or existing clients for possible intervention. Data collected in such population

screens, particularly when tied to any clinical progression to conditions like cancer, will be of value in the operations of, for example, health maintenance organizations, public health programs and insurance companies. Such data arrays or collections can be stored in machine-readable media and used in any number of health-related data management systems to provide improved healthcare services, cost effective healthcare, improved insurance operation, etc. See, for example, U.S. Patent Application No. 2002/0038227; U.S. Patent Application No. US 2004/0122296; U.S. Patent Application No. US 2004/ 0122297; and U.S. Patent No. 5,018,067. Such systems can access the data directly from internal data storage or remotely from one or more data storage sites as further detailed herein.

Each program can be implemented in a high level procedural or object oriented programming language to communicate with a computer system. However, the programs can be implemented in assembly or machine language, if desired. The language can be a compiled or interpreted language. Each such computer program can be stored on a storage media or device (e.g., ROM or magnetic diskette or others as defined elsewhere in this disclosure) readable by a general or special purpose programmable computer, for configuring and operating the computer when the storage media or device is read by the computer to perform the procedures described herein. The health-related data management system of the invention may also be considered to be implemented as a computer-readable storage medium, configured with a computer program, where the storage medium so configured causes a computer to operate in a specific and predefined manner to perform various functions described herein.

Differences in the genetic makeup of subjects can result in differences in their relative abilities to metabolize various drugs, which may modulate the symptoms or risk factors of cancer or metastatic events. Subjects that have cancer, or at risk for developing cancer or a metastatic event can vary in age, ethnicity, and other parameters. Accordingly, detection of the EGFR G724S, G719S or L858R mutation disclosed herein, both alone and together in combination with known genetic factors for drug metabolism, allow for a pre-determined level of predictability that a putative therapeutic or prophylactic to be tested in a selected subject will be suitable for treating cancer in the subject.

Performance And Accuracy Measures Of The Invention

The performance and thus absolute and relative clinical usefulness of the invention may be assessed in multiple ways as noted above. Amongst the various assessments of performance, the invention is intended to provide accuracy in clinical diagnosis and prognosis. The accuracy of a diagnostic, predictive, or prognostic test, assay, or method concerns the ability of the test, assay, or method to distinguish between subjects responsive to chemotherapeutic treatment and those that are not, is based on whether the subjects have the EGFR G724S, G719S, L858R mutation, or any combination thereof.

In the categorical diagnosis of a disease state, changing the cut point or threshold value of a test (or assay) usually changes the sensitivity and specificity, but in a qualitatively inverse relationship. Therefore, in assessing the accuracy and usefulness of a proposed medical test, assay, or method for assessing a subject's condition, one should always take both sensitivity and specificity into account and be mindful of what the cut point is at which the sensitivity and specificity are being reported because sensitivity and specificity may vary significantly over the range of cut points. Use of statistics such as AUC, encompassing all potential cut point values, is preferred for most categorical risk measures using the invention, while for continuous risk measures, statistics of goodness-of-fit and calibration to observed results or other gold standards, are preferred.

Using such statistics, an "acceptable degree of diagnostic accuracy", is herein defined as a test or assay in which the AUC (area under the ROC curve for the test or assay) is at least 0.60, desirably at least 0.65, more desirably at least 0.70, preferably at least 0.75, more preferably at least 0.80, and most preferably at least 0.85.

By a "very high degree of diagnostic accuracy", it is meant a test or assay in which the AUC (area under the ROC curve for the test or assay) is at least 0.80, desirably at least 0.85, more desirably at least 0.875, preferably at least 0.90, more preferably at least 0.925, and most preferably at least 0.95.

The predictive value of any test depends on the sensitivity and specificity of the test, and on the prevalence of the condition in the population being tested. This notion, based on Bayes' theorem, provides that the greater the likelihood that the condition being screened for is present in an individual or in the population (pre-test probability), the greater the validity of a positive test and the greater the likelihood that the result is a true positive. Thus, the problem with using a test in any population where there is a low likelihood of the condition

being present is that a positive result has limited value (i.e., more likely to be a false positive). Similarly, in populations at very high risk, a negative test result is more likely to be a false negative.

As a result, ROC and AUC can be misleading as to the clinical utility of a test in low
5 disease prevalence tested populations (defined as those with less than 1% rate of occurrences (incidence) per annum, or less than 10% cumulative prevalence over a specified time horizon). Alternatively, absolute risk and relative risk ratios as defined elsewhere in this disclosure can be employed to determine the degree of clinical utility. Populations of subjects to be tested can also be categorized into quartiles by the test's measurement values,
10 where the top quartile (25% of the population) comprises the group of subjects with the highest relative risk for therapeutic unresponsiveness, and the bottom quartile comprising the group of subjects having the lowest relative risk for therapeutic unresponsiveness. Generally, values derived from tests or assays having over 2.5 times the relative risk from top to bottom quartile in a low prevalence population are considered to have a "high degree of diagnostic
15 accuracy," and those with five to seven times the relative risk for each quartile are considered to have a "very high degree of diagnostic accuracy." Nonetheless, values derived from tests or assays having only 1.2 to 2.5 times the relative risk for each quartile remain clinically useful are widely used as risk factors for a disease; such is the case with total cholesterol and for many inflammatory biomarkers with respect to their prediction of future events. Often
20 such lower diagnostic accuracy tests must be combined with additional parameters in order to derive meaningful clinical thresholds for therapeutic intervention, as is done with the aforementioned global risk assessment indices.

A health economic utility function is yet another means of measuring the performance and clinical value of a given test, consisting of weighting the potential categorical test
25 outcomes based on actual measures of clinical and economic value for each. Health economic performance is closely related to accuracy, as a health economic utility function specifically assigns an economic value for the benefits of correct classification and the costs of misclassification of tested subjects. As a performance measure, it is not unusual to require a test to achieve a level of performance which results in an increase in health economic value
30 per test (prior to testing costs) in excess of the target price of the test.

In general, alternative methods of determining diagnostic accuracy are commonly used for continuous measures, when a disease category or risk category has not yet been

clearly defined by the relevant medical societies and practice of medicine, where thresholds for therapeutic use are not yet established, or where there is no existing gold standard for diagnosis of the pre-disease. For continuous measures of risk, measures of diagnostic accuracy for a calculated index are typically based on curve fit and calibration between the predicted continuous value and the actual observed values (or a historical index calculated value) and utilize measures such as R squared, Hosmer- Lemeshow *P*-value statistics and confidence intervals. It is not unusual for predicted values using such algorithms to be reported including a confidence interval (usually 90% or 95% CI) based on a historical observed cohort's predictions, as in the test for risk of future breast cancer recurrence commercialized by Genomic Health, Inc. (Redwood City, California).

DETECTION OF THE EGFR MUTATIONS

The actual detection of the EGFR G724S, G719S or L858R mutation can be determined at the protein or nucleic acid level using any method known in the art.

Wild type human EGFR nucleic acids and polypeptides have previously been described. For example, GenBank accession No: CAA25240.1, GenBank accession No: X00588.1, each of which is incorporated herein by reference in its entirety. A skilled artisan in the field would easily identify the wild type human EGFR nucleic acids and polypeptides sequences.

Human EGFR amino acid sequence (GenBank: CAA25240.1)

1	mrpsgtagaa llallaalcp asraleekkv cqgtsnkltq lgtfedhfls lqrmfnncev
61	vlgnleityv qrnydlsflk tiqevagyvl ialntverip lenlqiirgn myyensyala
121	vlsnydankt glkelpmrnl qeilhgavrf snnpalcnve siqwrdivss dflsnmsmdf
181	qnhlgscqkc dpscpngscw gageencqkl tkiicaqqcs grcrgkspsd cchnqcaagc
241	tgpresdclv crkfrdeatc kdtcpplmly npttyqmdvn pegkysfgat cvkkcprnyv
301	vtdhgscvra cgadsyemee dgvrkckkce gpcrkvcngi gigefkdsls inatnikhfk
361	nctsisgdlh ilpvafrgds fthtppldpq eldilktvke itgflliqaw penrtdlhaf
421	enleiirgrt kqhgqfslav vslnitslgl rslkeisdgd viisgnknlc yantinwkk1
481	fgtsgqktki isnrgensck atgqvchalc spegcwgpep rdcvscrnvs rgrecvdkck
541	llegeprefv enseciqchp eclpqamnit ctgrgpdnci qcahyidgph cvktcpagvm
601	genntlvwky adaghvchlc hpnctygctg pglegcptng pkipsiatgm vgalllllllv
661	algiglfmrr rhivrkrtrlr rllqerelve pltpsgeapn qallrilket efkkikvlgs
721	gafgtvykg1 wipegekvi pvaikelrea tspkankeil deayvmavsd nphvcrl1gi
781	cltstvqlit qlmpfgclld yvrehkdnig sqyllnwcvg iakgmnyled rrlvhrdlaa
841	rnlv1ktpqh vkitdfglak llgaeekyeh aeggkvpikw malesilhri ythqsdvwsy
901	gvtvwelmtf gskpydgipa seissilekg erlpqpict idvymimvkc wmidadsrpk

```

961  freliiefsk mardpqrylv iqqdermhlp sptdsnfyra lmdeedmddv vdadeylipq
1021 qgffsspsts rtpllsslsa tsnnstvaci drnglqscpi kedsflqrys sdptgalted
1081 siddtflvpv eyinqsvpkr pagsvqnpvy hnqplnpaps rdphyqdphs tavgnpeyln
1141 tvqptcvnst fdspahwaqk gshqislndp dyqqdffpke akpngifkgs taenaeylrv
1201 apqssefiga

```

Human EGFR nucleic acid sequence (GenBank: X00588.1)

```

1  gccgcgctgc gccggagtcc cgagctagcc ccggcgccgc cgccgcccag accggacgac
61  aggccacctc gtcggcgctcc gcccgagtcc ccgcctcgcc gccaacgcca caaccaccgc
121 gcacggcccc ctgactccgt ccagtattga tcgggagagc cggagcgagc tcttcgggga
181 gcagcgatgc gaccctccgg gacggccggg gcagcgctcc tggcgctgct ggctgcgctc
241 tgcccggcga gtcgggctct ggaggaaaag aaagtttgcc aaggcacgag taacaagctc
301 acgcagttgg gcacttttga agatcatttt ctcagcctcc agaggatgtt caataactgt
361 gaggtggtcc ttgggaatth ggaaattacc tatgtgcaga ggaattatga tctttccttc
421 ttaaagacca tccaggaggt ggctggttat gtctcattg ccctcaacac agtggagcga
481 attcctttgg aaaacctgca gatcatcaga ggaaatatgt actacgaaaa ttcctatgcc
541 ttagcagtct tatctaacta tgatgcaaat aaaaccggac tgaaggagct gcccatgaga
601 aatttacagg aaatcctgca tggcgccgtg cggttcagca acaaccctgc cctgtgcaac
661 gtggagagca tccagtggcg ggacatagtc agcagtgact ttctcagcaa catgtcgatg
721 gacttccaga accacctggg cagctgccaa aagtgtgatc caagctgtcc caatgggagc
781 tgctggggtg caggagagga gaactgccag aaactgacca aaatcatctg tgcccagcag
841 tgctccgggc gctgccgtgg caagtcccc agtgactgct gccacaacca gtgtgctgca
901 ggctgcacag gccccggga gagcgactgc ctggtctgcc gcaaattccg agacgaagcc
961 acgtgcaagg acacctgccc cccactcatg ctctacaacc ccaccacgta ccagatggat
1021 gtgaaccccg agggcaaata cagctttggt gccacctgcg tgaagaagtg tccccgtaat
1081 tatgtggtga cagatcacgg ctctgtcgctc cgagcctgtg gggccgacag ctatgagatg
1141 gaggaagacg gcgtccgcaa gtgtaagaag tgcgaagggc cttgccgcaa agtgtgtaac
1201 ggaataggta ttggtgaatt taaagactca ctctccataa atgctacgaa tattaacac
1261 ttcaaaaact gcacctccat cagtggcgat ctccacatcc tgccggtggc atttaggggt
1321 gactccttca cacatactcc tcctctggat ccacaggaac tggatattct gaaaaccgta
1381 aaggaaatca cagggttttt gctgattcag gcttggcctg aaaacaggac ggacctccat
1441 gcctttgaga acctagaaat catacgcggc aggaccaagc aacatggtca gttttctctt
1501 gcagtcgtca gcctgaacat aacatccttg ggattacgct ccctcaagga gataagtgat
1561 ggagatgtga taatttcagg aaacaaaaat ttgtgctatg caaatacaat aaactggaaa
1621 aaactgtttg ggacctccgg tcagaaaacc aaaattataa gcaacagagg tgaaaacagc
1681 tgcaaggcca caggccaggt ctgccatgcc ttgtgctccc ccgagggctg ctggggcccg
1741 gagcccaggg actgcgctctc ttgccggaat gtcagccgag gcagggaatg cgtggacaag
1801 tgcaagcttc tggaggggtg gccaaaggag tttgtggaga actctgagtg catacagtgc
1861 caccagaggt gcctgcctca ggccatgaac atcacctgca caggacgggg accagacaac
1921 tgtatccagt gtgcccacta cattgacggc cccactgcg tcaagacctg cccggcagga
1981 gtcattgggag aaaacaacac cctggtctgg aagtacgcag acgccggcca tgtgtgccac

```

```

2041 ctgtgccatc caaactgcac ctacggatgc actgggccag gtcttgaagg ctgtccaacg
2101 aatgggccta agatcccgtc catcgccact gggatggtgg gggccctcct cttgctgctg
2161 gtggtggccc tggggatcgg cctcttcacg cgaaggcgcc acatcgttcg gaagcgacg
2221 ctgcggaggg tgctgcagga gagggagctt gtggagcctc ttacaccagc tggagaagct
2281 cccaaccaag ctctcttgag gatcttgaag gaaactgaat tcaaaaagat caaagtgctg
2341 ggctccggtg cgttcggcac ggtgtataag ggactctgga tcccagaagg tgagaaagtt
2401 aaaattcccg tcgctatcaa ggaattaaga gaagcaacat ctccgaaagc caacaaggaa
2461 atcctcgatg aagcctacgt gatggccagc gtggacaacc cccacgtgtg ccgctgctg
2521 ggcattctgc tcacctccac cgtgcaactc atcacgcagc tcatgccctt cggctgcctc
2581 ctggactatg tccgggaaca caaagacaat attggctccc agtacctgct caactgggtg
2641 gtgcagatcg caaagggcat gaactacttg gaggaccgtc gcttggtgca ccgcgacctg
2701 gcagccagga acgtactggt gaaaacaccg cagcatgtca agatcacaga ttttgggctg
2761 gccaaactgc tgggtgcgga agagaaagaa taccatgcag aaggaggcaa agtgcctatc
2821 aagtggatgg cattggaatc aattttacac agaattctata cccaccagag tgatgtctgg
2881 agctacgggg tgaccgtttg ggagttgatg acctttggat ccaagccata tgacggaatc
2941 cctgccagcg agatctcctc catcctggag aaaggagaac gcctccctca gccaccata
3001 tgtaccatcg atgtctacat gatcatggtc aagtgtgga tgatagacgc agatagtgcg
3061 ccaaagttcc gtgagttgat catcgaattc tccaaaatgg cccgagaccc ccagcgctac
3121 cttgtcattc agggggatga aagaatgcat ttgccaaagtc ctacagactc caacttctac
3181 cgtgccctga tggatgaaga agacatggac gacgtggtgg atgccgacga gtacctcatc
3241 ccacagcagg gcttcttcag cagcccctcc acgtcacgga ctcccctcct gagctctctg
3301 agtgcaacca gcaacaattc caccgtggct tgcattgata gaaatgggct gcaaagctgt
3361 cccatcaagg aagacagctt cttgcagcga tacagctcag accccacagg cgccttgact
3421 gaggacagca tagacgacac cttcctccca gtgcctgaat acataaacca gtccgttccc
3481 aaaaggcccg ctggctctgt gcagaatcct gtctatcaca atcagcctct gaaccccgcg
3541 cccagcagag acccacacta ccaggacccc cacagcactg cagtgggcaa ccccgagtat
3601 ctcaacactg tccagcccac ctgtgtcaac agcacattcg acagccctgc ccactgggcc
3661 cagaaaggca gccaccaaat tagcctggac aaccctgact accagcagga cttctttccc
3721 aagggaagcca agccaaatgg catctttaag ggctccacag ctgaaaatgc agaataccta
3781 agggtcgctg cacaaagcag tgaatttatt ggagcatgac cacggaggat agtatgagcc
3841 ctaaaaatcc agactctttc gatacccagg accaagccac agcaggctct ccatcccaac
3901 agccatgccc gcattagctc ttagaccac agactggttt tgcaacgttt acaccgacta
3961 gccaggaagt acttccacct cgggcacatt ttgggaagtt gcattccttt gtcttcaaac
4021 tgtgaagcat ttacagaaac gcatccagca agaattattgt ccctttgagc agaaatttat
4081 ctttcaaaga ggtatatattg aaaaaaaaaa aaaaagtata tgtgaggatt tttattgatt
4141 ggggatcttg gagtttttca ttgtcgctat tgatttttac ttcaatgggc tcttccaaca
4201 aggaagaagc ttgctggtag cacttgctac cctgagttca tccaggccca actgtgagca
4261 aggagcacia gccacaagtc ttccagagga tgcttgattc cagtggttct gcttcaaggc
4321 ttccactgca aaacactaaa gatccaagaa ggcttcatg gcccagcag gccggatcgg
4381 tactgtatca agtcatggca ggtacagtag gataagccac tctgtccctt cctgggcaaa
4441 gaagaaacgg aggggatgaa ttcttcctta gacttacttt tgtaaaaatg tccccacggt
4501 acttactccc cactgatgga ccagtggttt ccagtcatga gcgttagact gacttgtttg
4561 tcttccattc cattgttttg aaactcagta tgccgccctt gtcttgctgt catgaaatca

```

```

4621 gcaagagagg atgacacatc aaataataac tcggattcca gcccacattg gattcatcag
4681 cttttggacc aatagcccac agctgagaat gtggaatacc taaggataac accgcttttg
4741 ttctcgcaaa aacgtatctc ctaatttgag gctcagatga aatgcatcag gtcctttggg
4801 gcatagatca gaagactaca aaaatgaagc tgctctgaaa tctcctttag ccatcacccc
4861 aaccccccaa aattagtttg tgttacttat ggaagatagt tttctccttt tacttcactt
4921 caaaagcttt ttactcaaag agtatatgtt ccctccaggt cagctgcccc caaacccctt
4981 ccttacgctt tgtcacacaa aaagtgtctc tgcttgagt catctattca agcacttaca
5041 gctctggcca caacagggca ttttacaggt gcgaatgaca gtagcattat gagtagtgtg
5101 aattcaggta gtaaatatga aactagggtt tgaaattgat aatgctttca caacatttgc
5161 agatgtttta gaaggaaaaa agttccttcc taaaataatt tctctacaat tggaagattg
5221 gaagattcag ctagttagga gcccatTTTT tcctaactctg tgtgtgccct gtaacctgac
5281 tggttaacag cagtcctttg taaacagtgt tttaaactct cctagtcaat atccacccca
5341 tccaatttat caaggaagaa atggttcaga aaatattttc agcctacagt tatgttcagt
5401 cacacacaca tacaaaatgt tccttttgct tttaaagtaa tttttgactc ccagatcagt
5461 cagagcccct acagcattgt taagaaagta tttgattttt gtctcaatga aaataaaact
5521 atattcattt cc

```

G724S, G719S or L858R mutation -specific reagents useful in the practice of the disclosed methods include, among others, mutant polypeptide specific antibodies and AQUA peptides (heavy-isotope labeled peptides) corresponding to, and suitable for detection and quantification of, mutant (G724S, G719S, L858R, or any combination thereof) polypeptide expression in a biological sample. A mutant polypeptide-specific reagent is any reagent, biological or chemical, capable of specifically binding to, detecting and/or quantifying the presence/level of expressed mutant polypeptide in a biological sample, while not binding to or detecting wild type. The term includes, but is not limited to, the preferred antibody and AQUA peptide reagents discussed below, and equivalent reagents are within the scope of the present invention.

Reagents suitable for use in practice of the methods of the invention include a mutant polypeptide-specific antibody. A mutant-specific antibody of the invention is an isolated antibody or antibodies that specifically bind(s) a mutant (G724S, G719S, L858R, or any combination thereof) polypeptide of the invention, but does not substantially bind either wild type or when mutated at other positions.

Mutant polypeptide-specific antibodies generated against human mutant may also bind to highly homologous and equivalent epitopic peptide sequences in other mammalian species, for example murine, rat, feline, pig, or rabbit, and vice versa. Antibodies useful in practicing the methods of the invention include (a) monoclonal antibodies, (b) purified

polyclonal antibodies that specifically bind to the target polypeptide (e.g. an epitope comprising the G724S, G719S, L858R mutation point, or any combination thereof), (c) antibodies as described in (a)-(b) above that bind equivalent and highly homologous epitopes or phosphorylation sites in other non-human species (e.g. mouse, rat), and (d) fragments of (a)-(c) above that bind to the antigen (or more preferably the epitope) bound by the exemplary antibodies disclosed herein.

The term "antibody" or "antibodies" as used herein refers to all types of immunoglobulins, including IgG, IgM, IgA, IgD, and IgE. The antibodies may be monoclonal or polyclonal and may be of any species of origin, including (for example) mouse, rat, rabbit, horse, or human, or may be chimeric antibodies. See, e.g., M. Walker et al., *Molec. Immunol.* 26: 403-11 (1989); Morrisson et al., *Proc. Nat'l. Acad. Sci.* 81:6851 (1984); Neuberger et al., *Nature* 312:604 (1984)). The antibodies may be recombinant monoclonal antibodies produced according to the methods disclosed in U.S. Pat. No. 4,474,893 (Reading) or U.S. Pat. No. 4,816,567 (Cabilly et al.) The antibodies may also be chemically constructed specific antibodies made according to the method disclosed in U.S. Pat. No. 4,676,980 (Segel et al.)

The invention is not limited to use of antibodies, but includes equivalent molecules, such as protein binding domains or nucleic acid aptamers, which bind, in a mutant-protein or truncated-protein specific manner, to essentially the same epitope to which a mutant polypeptide-specific antibody useful in the methods of the invention binds. See, e.g., Neuberger et al., *Nature* 312: 604 (1984). Such equivalent non-antibody reagents may be suitably employed in the methods of the invention further described below.

Polyclonal antibodies useful in practicing the methods of the invention may be produced according to standard techniques by immunizing a suitable animal (e.g., rabbit, goat, etc.) with an antigen encompassing a desired mutant-protein specific epitope (e.g. the sequence comprising the G724S, G719S, L858R mutation site, or any combination thereof) collecting immune serum from the animal, and separating the polyclonal antibodies from the immune serum, and purifying polyclonal antibodies having the desired specificity, in accordance with known procedures. The antigen may be a synthetic peptide antigen comprising the desired epitopic sequence, selected and constructed in accordance with well-known techniques. See, e.g., *ANTIBODIES: A LABORATORY MANUAL*, Chapter 5, p. 75-76, Harlow & Lane Eds., Cold Spring Harbor Laboratory (1988); Czernik, *Methods In*

Enzymology, 201:264-283 (1991); Merrifield, J. Am. Chem. Soc. 85: 21-49 (1962)).

Polyclonal antibodies produced as described herein may be screened and isolated as further described below.

Monoclonal antibodies may also be beneficially employed in the methods of the invention, and may be produced in hybridoma cell lines according to the well-known technique of Kohler and Milstein. Nature 265:495-97 (1975); Kohler and Milstein, Eur. J. Immunol. 6: 511 (1976); see also, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Ausubel et al., Eds. (1989). Monoclonal antibodies so produced are highly specific, and improve the selectivity and specificity of assay methods provided by the invention. For example, a solution containing the appropriate antigen (e.g. a synthetic peptide comprising the mutant junction of mutant EGFR polypeptide) may be injected into a mouse and, after a sufficient time (in keeping with conventional techniques), the mouse sacrificed and spleen cells obtained. The spleen cells are then immortalized by fusing them with myeloma cells, typically in the presence of polyethylene glycol, to produce hybridoma cells. Rabbit mutant hybridomas, for example, may be produced as described in U.S. Pat. No. 5,675,063, K. Knight, Issued Oct. 7, 1997. The hybridoma cells are then grown in a suitable selection media, such as hypoxanthine-aminopterin-thymidine (HAT), and the supernatant screened for monoclonal antibodies having the desired specificity, as described below. The secreted antibody may be recovered from tissue culture supernatant by conventional methods such as precipitation, ion exchange or affinity chromatography, or the like.

Monoclonal Fab fragments may also be produced in Escherichia coli by recombinant techniques known to those skilled in the art. See, e.g., W. Huse, Science 246:1275-81 (1989); Mullinax et al., Proc. Nat'l Acad. Sci. 87: 8095 (1990). If monoclonal antibodies of one isotype are preferred for a particular application, particular isotypes can be prepared directly, by selecting from the initial mutant, or prepared secondarily, from a parental hybridoma secreting a monoclonal antibody of different isotype by using the sib selection technique to isolate class-switch variants (Steplewski, et al., Proc. Nat'l. Acad. Sci., 82:8653 (1985); Spira et al., J. Immunol. Methods, 74:307 (1984)). The antigen combining site of the monoclonal antibody can be cloned by PCR and single-chain antibodies produced as phage-displayed recombinant antibodies or soluble antibodies in E. coli (see, e.g., ANTIBODY ENGINEERING PROTOCOLS, 1995, Humana Press, Sudhir Paul editor.)

Further still, U.S. Pat. No. 5,194,392, Geysen (1990) describes a general method of

detecting or determining the sequence of monomers (amino acids or other compounds) which is a topological equivalent of the epitope (i.e., a "mimotope") which is complementary to a particular paratope (antigen binding site) of an antibody of interest. More generally, this method involves detecting or determining a sequence of monomers which is a topographical equivalent of a ligand which is complementary to the ligand binding site of a particular receptor of interest. Similarly, U.S. Pat. No. 5,480,971, Houghten et al., (1996) discloses linear C.sub.1-C-alkyl peralkylated oligopeptides and sets and libraries of such peptides, as well as methods for using such oligopeptide sets and libraries for determining the sequence of a peralkylated oligopeptide that preferentially binds to an acceptor molecule of interest. Thus, non-peptide analogs of the epitope-bearing peptides of the invention also can be made routinely by these methods.

Antibodies useful in the methods of the invention, whether polyclonal or monoclonal, may be screened for epitope and mutant protein specificity according to standard techniques. See, e.g. Czernik et al., *Methods in Enzymology*, 201: 264-283 (1991). For example, the antibodies may be screened against a peptide library by ELISA to ensure specificity for both the desired antigen and, if desired, for reactivity only with a mutant polypeptide of the invention and not with wild type. The antibodies may also be tested by Western blotting against cell preparations containing target protein to confirm reactivity with the only the desired target and to ensure no appreciable binding to other mutants not containing the G724S point mutation. The production, screening, and use of mutant protein-specific antibodies is known to those of skill in the art, and has been described.

Mutant polypeptide-specific antibodies useful in the methods of the invention may exhibit some limited cross-reactivity with similar epitopes in other highly homologous proteins. This is not unexpected as most antibodies exhibit some degree of cross-reactivity, and anti-peptide antibodies will often cross-react with epitopes having high homology or identity to the immunizing peptide. See, e.g., Czernik, *supra*. Cross-reactivity with other mutant proteins is readily characterized by Western blotting alongside markers of known molecular weight. Amino acid sequences of cross-reacting proteins may be examined to identify sites highly homologous or identical to the mutant polypeptide sequence to which the antibody binds. Undesirable cross-reactivity can be removed by negative selection using antibody purification on peptide columns (e.g. selecting out antibodies that bind either wild type EGFR polypeptide or which bind highly homologous sequences on different proteins).

Mutant EGFR polypeptide-specific antibodies of the invention that are useful in practicing the methods disclosed herein are ideally specific for human mutant polypeptide, but are not limited only to binding the human species, per se. The invention includes the production and use of antibodies that also bind conserved and highly homologous or identical epitopes in other mammalian species (e.g. mouse, rat, monkey). Highly homologous or identical sequences in other species can readily be identified by standard sequence comparisons, such as using BLAST, with the human mutant EGFR polypeptide sequence the proteins sequences of EGFR from other mammalian species have been published and are available in the SwissProt database.

Antibodies employed in the methods of the invention may be further characterized by, and validated for, use in a particular assay format, for example FC, IHC, and/or ICC. The use of mutant EGFR polypeptide-specific antibodies in such methods is further described below. Antibodies may also be advantageously conjugated to fluorescent dyes (e.g. Alexa488, PE), or labels such as quantum dots, for use in multi-parametric analyses along with other signal transduction (phospho-AKT, phospho-Erk 1/2) and/or cell marker (cytokeratin) antibodies, as further described below.

In practicing the methods of the invention, the expression and/or activity of wild type EGFR in a given biological sample may also be advantageously examined using antibodies (either phospho-specific or total) for these wild type proteins. Such antibodies may also be produced according to standard methods, as described above. The amino acid sequences of human EGFR are published, as are the sequences of these proteins from other species, including mammalian, as noted above.

Detection of wild type EGFR expression and/or activation, along with mutant EGFR polypeptide expression, in a biological sample (e.g. a tumor sample) can provide information on whether the mutant protein alone is driving the tumor, or whether wild type EGFR is also activated and driving the tumor. Such information is clinically useful in assessing whether targeting the mutant protein or the wild type protein(s), or both, or is likely to be most beneficial in inhibiting progression of the tumor, and in selecting an appropriate therapeutic or combination thereof.

It will be understood that more than one antibody may be used in the practice of the above-described methods. For example, one or more mutant EGFR polypeptide-specific antibodies together with one or more antibodies specific for another protein, receptor, that is

suspected of being, or potentially is, activated in a cancer in which mutant EGFR polypeptide is expressed may be simultaneously employed to detect the activity of such other signaling molecules in a biological sample comprising cells from such cancer.

Those of skill in the art will appreciate that mutant EGFR polypeptides of the present invention and the mutant junction epitope-bearing fragments thereof described above can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric polypeptides. These mutant proteins facilitate purification and show an increased half-life in vivo. This has been shown, e.g., for chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins (EPA 394,827; Traunecker et al., Nature 331: 84-86 (1988)). Mutant proteins that have a disulfide-linked dimeric structure due to the IgG part can also be more efficient in binding and neutralizing other molecules than the monomeric mutant EGFR polypeptide alone (Fountoulakis et al., J Biochem 270: 3958-3964 (1995)).

Mutant EGFR polypeptide-specific reagents useful in the practice of the disclosed methods may also comprise heavy-isotope labeled peptides suitable for the absolute quantification of expressed mutant EGFR polypeptide in a biological sample. The production and use of AQUA peptides for the absolute quantification of proteins (AQUA) in complex mixtures has been described. See WO/03016861, "Absolute Quantification of Proteins and Modified Forms Thereof by Multistage Mass Spectrometry," Gygi et al. and also Gerber et al., Proc. Natl. Acad. Sci. U.S.A. 100: 6940-5 (2003) (the teachings of which are hereby incorporated herein by reference, in their entirety).

The AQUA methodology employs the introduction of a known quantity of at least one heavy-isotope labeled peptide standard (which has a unique signature detectable by LC-SRM chromatography) into a digested biological sample in order to determine, by comparison to the peptide standard, the absolute quantity of a peptide with the same sequence and protein modification in the biological sample. Briefly, the AQUA methodology has two stages: peptide internal standard selection and validation and method development; and implementation using validated peptide internal standards to detect and quantify a target protein in sample. The method is a powerful technique for detecting and quantifying a given peptide/protein within a complex biological mixture, such as a cell lysate, and may be employed, e.g., to quantify change in protein phosphorylation as a result of drug treatment, or to quantify differences in the level of a protein in different biological states.

Generally, to develop a suitable internal standard, a particular peptide (or modified peptide) within a target protein sequence is chosen based on its amino acid sequence and the particular protease to be used to digest. The peptide is then generated by solid-phase peptide synthesis such that one residue is replaced with that same residue containing stable isotopes

5 The result is a peptide that is chemically identical to its native counterpart formed by proteolysis, but is easily distinguishable by MS via a 7-Da mass shift. The newly synthesized AQUA internal standard peptide is then evaluated by LC-MS/MS. This process provides qualitative information about peptide retention by reverse-phase chromatography, ionization efficiency, and fragmentation via collision-induced dissociation. Informative and abundant
10 fragment ions for sets of native and internal standard peptides are chosen and then specifically monitored in rapid succession as a function of chromatographic retention to form a selected reaction monitoring (LC-SRM) method based on the unique profile of the peptide standard.

The second stage of the AQUA strategy is its implementation to measure the amount
15 of a protein or modified protein from complex mixtures. Whole cell lysates are typically fractionated by SDS-PAGE gel electrophoresis, and regions of the gel consistent with protein migration are excised. This process is followed by in-gel proteolysis in the presence of the AQUA peptides and LC-SRM analysis. (See Gerber et al. supra.) AQUA peptides are spiked in to the complex peptide mixture obtained by digestion of the whole cell lysate with a
20 proteolytic enzyme and subjected to immunoaffinity purification as described above. The retention time and fragmentation pattern of the native peptide formed by digestion (e.g. trypsinization) is identical to that of the AQUA internal standard peptide determined previously; thus, LC-MS/MS analysis using an SRM experiment results in the highly specific and sensitive measurement of both internal standard and analyte directly from extremely
25 complex peptide mixtures.

Since an absolute amount of the AQUA peptide is added (*e.g.*, 250 fmol), the ratio of the areas under the curve can be used to determine the precise expression levels of a protein or phosphorylated form of a protein in the original cell lysate. In addition, the internal standard is present during in-gel digestion as native peptides are formed, such that peptide
30 extraction efficiency from gel pieces, absolute losses during sample handling (including vacuum centrifugation), and variability during introduction into the LC-MS system do not affect the determined ratio of native and AQUA peptide abundances.

An AQUA peptide standard is developed for a known sequence previously identified by the IAP-LC-MS/MS method within in a target protein. If the site is modified, one AQUA peptide incorporating the modified form of the particular residue within the site may be developed, and a second AQUA peptide incorporating the unmodified form of the residue developed. In this way, the two standards may be used to detect and quantify both the modified an unmodified forms of the site in a biological sample.

Peptide internal standards may also be generated by examining the primary amino acid sequence of a protein and determining the boundaries of peptides produced by protease cleavage. Alternatively, a protein may actually be digested with a protease and a particular peptide fragment produced can then sequenced. Suitable proteases include, but are not limited to, serine proteases (e.g. trypsin, hepsin), metallo proteases (e.g. PUMP1), chymotrypsin, cathepsin, pepsin, thermolysin, carboxypeptidases, etc.

A peptide sequence within a target protein is selected according to one or more criteria to optimize the use of the peptide as an internal standard. Preferably, the size of the peptide is selected to minimize the chances that the peptide sequence will be repeated elsewhere in other non-target proteins. Thus, a peptide is preferably at least about 6 amino acids. The size of the peptide is also optimized to maximize ionization frequency. Thus, peptides longer than about 20 amino acids are not preferred. The preferred ranged is about 7 to 15 amino acids. A peptide sequence is also selected that is not likely to be chemically reactive during mass spectrometry, thus sequences comprising cysteine, tryptophan, or methionine are avoided.

A peptide sequence that does not include a modified region of the target region may be selected so that the peptide internal standard can be used to determine the quantity of all forms of the protein. Alternatively, a peptide internal standard encompassing a modified amino acid may be desirable to detect and quantify only the modified form of the target protein. Peptide standards for both modified and unmodified regions can be used together, to determine the extent of a modification in a particular sample (i.e. to determine what fraction of the total amount of protein is represented by the modified form). For example, peptide standards for both the phosphorylated and unphosphorylated form of a protein known to be phosphorylated at a particular site can be used to quantify the amount of phosphorylated form in a sample.

The peptide is labeled using one or more labeled amino acids (i.e. the label is an actual part of the peptide) or less preferably, labels may be attached after synthesis according to standard methods. Preferably, the label is a mass-altering label selected based on the following considerations: The mass should be unique to shift fragments masses produced by MS analysis to regions of the spectrum with low background; the ion mass signature component is the portion of the labeling moiety that preferably exhibits a unique ion mass signature in MS analysis; the sum of the masses of the constituent atoms of the label is preferably uniquely different than the fragments of all the possible amino acids. As a result, the labeled amino acids and peptides are readily distinguished from unlabeled ones by the ion/mass pattern in the resulting mass spectrum. Preferably, the ion mass signature component imparts a mass to a protein fragment that does not match the residue mass for any of the 20 natural amino acids.

The label should be robust under the fragmentation conditions of MS and not undergo unfavorable fragmentation. Labeling chemistry should be efficient under a range of conditions, particularly denaturing conditions, and the labeled tag preferably remains soluble in the MS buffer system of choice. The label preferably does not suppress the ionization efficiency of the protein and is not chemically reactive. The label may contain a mixture of two or more isotopically distinct species to generate a unique mass spectrometric pattern at each labeled fragment position. Stable isotopes, such as ^2H , ^{13}C , ^{15}N , ^{17}O , ^{18}O , or ^{34}S , are among preferred labels. Pairs of peptide internal standards that incorporate a different isotope label may also be prepared. Preferred amino acid residues into which a heavy isotope label may be incorporated include leucine, proline, valine, and phenylalanine.

Peptide internal standards are characterized according to their mass-to-charge (m/z) ratio, and preferably, also according to their retention time on a chromatographic column (e.g., an HPLC column). Internal standards that co-elute with unlabeled peptides of identical sequence are selected as optimal internal standards. The internal standard is then analyzed by fragmenting the peptide by any suitable means, for example by collision-induced dissociation (CID) using, e.g., argon or helium as a collision gas. The fragments are then analyzed, for example by multi-stage mass spectrometry (MS) to obtain a fragment ion spectrum, to obtain a peptide fragmentation signature. Preferably, peptide fragments have significant differences in m/z ratios to enable peaks corresponding to each fragment to be well separated, and a signature is that is unique for the target peptide is obtained. If a suitable fragment signature is

not obtained at the first stage, additional stages of MS are performed until a unique signature is obtained.

Fragment ions in the MS/MS and MS³ spectra are typically highly specific for the peptide of interest, and, in conjunction with LC methods, allow a highly selective means of detecting and quantifying a target peptide/protein in a complex protein mixture, such as a cell lysate, containing many thousands or tens of thousands of proteins. Any biological sample potentially containing a target protein/peptide of interest may be assayed. Crude or partially purified cell extracts are preferably employed. Generally, the sample has at least 0.01 mg of protein, typically a concentration of 0.1-10 mg/mL, and may be adjusted to a desired buffer concentration and pH.

A known amount of a labeled peptide internal standard, preferably about 10 femtomoles, corresponding to a target protein to be detected/quantified is then added to a biological sample, such as a cell lysate. The spiked sample is then digested with one or more protease(s) for a suitable time period to allow digestion. A separation is then performed (*e.g.* by HPLC, reverse-phase HPLC, capillary electrophoresis, ion exchange chromatography, etc.) to isolate the labeled internal standard and its corresponding target peptide from other peptides in the sample. Microcapillary LC is a preferred method.

Each isolated peptide is then examined by monitoring of a selected reaction in the MS. This involves using the prior knowledge gained by the characterization of the peptide internal standard and then requiring the MS to continuously monitor a specific ion in the MS/MS or MSⁿ spectrum for both the peptide of interest and the internal standard. After elution, the area under the curve (AUC) for both peptide standard and target peptide peaks are calculated. The ratio of the two areas provides the absolute quantification that can be normalized for the number of cells used in the analysis and the protein's molecular weight, to provide the precise number of copies of the protein per cell. Further details of the AQUA methodology are described in Gygi *et al.*, and Gerber *et al. supra.* AQUA internal peptide standards (heavy-isotope labeled peptides) may desirably be produced, as described above, to detect and quantify any unique site (*e.g.* the G724S mutation point) within a mutant EGFR polypeptide of the invention. For example, an AQUA phosphopeptide may be prepared that corresponds to the EGFR peptide sequence immediately encompassing the G724S mutation point. Peptide standards may be produced and such standards employed in the AQUA methodology to detect and quantify the presence of mutant (G724S, G719S, L858R, or any

combination thereof) EGFR (*i.e.* the presence of the peptide sequence encompassing the G724S, G719S, L858R point mutation, or any combination thereof) in a biological sample.

For example, an exemplary AQUA peptide of the invention comprises the amino acid sequence EVA, which corresponds to the three amino acids immediately flanking each side of the G724S, G719S or L858R mutation point in the mutant EGFR polypeptide. It will be appreciated that larger AQUA peptides comprising the mutant junction sequence (and additional residues downstream or upstream of it) may also be constructed. Similarly, a smaller AQUA peptide comprising less than all of the residues of such sequence (but still comprising the point of mutant junction itself) may alternatively be constructed. Such larger or shorter AQUA peptides are within the scope of the present invention, and the selection and production of preferred AQUA peptides may be carried out as described above (see Gygi et al., Gerber et al., *supra.*).

Mutant-specific reagents provided by the invention also include nucleic acid probes and primers suitable for detection of a mutant EGFR polynucleotide. The specific use of such probes in assays such as fluorescence in-situ hybridization (FISH) or polymerase chain reaction (PCR) amplification.

Also provided by the invention is a kit for the detection of mutant (G724S, G719S, L858R, or any combination thereof) EGFR in a biological sample, the kit comprising an isolated mutant EGFR-specific reagent of the invention and one or more secondary reagents. Suitable secondary reagents for employment in a kit are familiar to those of skill in the art, and include, by way of example, buffers, detectable secondary antibodies or probes, activating agents, and the like.

The methods of the invention may be carried out in a variety of different assay formats known to those of skill in the art.

Immunoassays useful in the practice of the methods of the invention may be homogenous immunoassays or heterogeneous immunoassays. In a homogeneous assay the immunological reaction usually involves a mutant EGFR polypeptide-specific reagent (e.g. a mutant EGFR polypeptide-specific antibody), a labeled analyte, and the biological sample of interest. The signal arising from the label is modified, directly or indirectly, upon the binding of the antibody to the labeled analyte. Both the immunological reaction and detection of the extent thereof are carried out in a homogeneous solution. Immunochemical labels that may be employed include free radicals, radio-isotopes, fluorescent dyes, enzymes, bacteriophages,

coenzymes, and so forth. Semi-conductor nanocrystal labels, or "quantum dots", may also be advantageously employed, and their preparation and use has been well described. See generally, K. Barovsky, *Nanotech. Law & Bus.* 1(2): Article 14 (2004) and patents cited therein.

5 In a heterogeneous assay approach, the reagents are usually the biological sample, a mutant EGFR polypeptide-specific reagent (e.g., a mutant EGFR-specific antibody), and suitable means for producing a detectable signal. Biological samples as further described below may be used. The antibody is generally immobilized on a support, such as a bead, plate or slide, and contacted with the sample suspected of containing the antigen in a liquid phase.
10 The support is then separated from the liquid phase and either the support phase or the liquid phase is examined for a detectable signal employing means for producing such signal. The signal is related to the presence of the analyte in the biological sample. Means for producing a detectable signal include the use of radioactive labels, fluorescent labels, enzyme labels, quantum dots, and so forth. For example, if the antigen to be detected contains a second
15 binding site, an antibody which binds to that site can be conjugated to a detectable group and added to the liquid phase reaction solution before the separation step. The presence of the detectable group on the solid support indicates the presence of the antigen in the test sample. Examples of suitable immunoassays are the radioimmunoassay, immunofluorescence methods, enzyme-linked immunoassays, and the like.

20 Immunoassay formats and variations thereof, which may be useful for carrying out the methods disclosed herein, are well known in the art. See generally E. Maggio, *Enzyme-Immunoassay*, (1980) (CRC Press, Inc., Boca Raton, Fla.); see also, e.g., U.S. Pat. No. 4,727,022 (Skold et al., "Methods for Modulating Ligand-Receptor Interactions and their Application"); U.S. Pat. No. 4,659,678 (Forrest et al., "Immunoassay of Antigens"); U.S. Pat.
25 No. 4,376,110 (David et al., "Immunometric Assays Using Monoclonal Antibodies"). Conditions suitable for the formation of reagent-antibody complexes are well known to those of skill in the art. See *id.* Mutant EGFR polypeptide-specific monoclonal antibodies may be used in a "two-site" or "sandwich" assay, with a single hybridoma cell line serving as a source for both the labeled monoclonal antibody and the bound monoclonal antibody. Such
30 assays are described in U.S. Pat. No. 4,376,110. The concentration of detectable reagent should be sufficient such that the binding of mutant EGFR polypeptide is detectable compared to background.

Antibodies useful in the practice of the methods disclosed herein may be conjugated to a solid support suitable for a diagnostic assay (e.g., beads, plates, slides or wells formed from materials such as latex or polystyrene) in accordance with known techniques, such as precipitation. Antibodies or other mutant EGFR polypeptide-binding reagents may likewise
5 be conjugated to detectable groups such as radiolabels (e.g., .sup.35S, .sup.125I, .sup.131I), enzyme labels (e.g., horseradish peroxidase, alkaline phosphatase), and fluorescent labels (e.g., fluorescein) in accordance with known techniques.

Cell-based assays, such flow cytometry (FC), immuno-histochemistry (IHC), or immunofluorescence (IF) are particularly desirable in practicing the methods of the invention,
10 since such assay formats are clinically-suitable, allow the detection of mutant EGFR polypeptide expression in vivo, and avoid the risk of artifact changes in activity resulting from manipulating cells obtained from, e.g. a tumor sample in order to obtain extracts. Accordingly, in some preferred embodiment, the methods of the invention are implemented in a flow-cytometry (FC), immuno-histochemistry (IHC), or immunofluorescence (IF) assay
15 format.

Flow cytometry (FC) may be employed to determine the expression of mutant EGFR polypeptide in a mammalian leukemia sample before, during, and after treatment with a drug targeted at inhibiting EGFR activity. For example, tumor cells from a bone marrow sample may be analyzed by flow cytometry for mutant EGFR polypeptide expression and/or
20 activation, as well as for markers identifying cancer cell types, etc., if so desired. Flow cytometry may be carried out according to standard methods. See, e.g. Chow et al., Cytometry (Communications in Clinical Cytometry) 46: 72-78 (2001). Briefly and by way of example, the following protocol for cytometric analysis may be employed: fixation of the cells with 2% paraformaldehyde for 10 minutes at 37.degree. C. followed by
25 permeabilization in 90% methanol for 30 minutes on ice. Cells may then be stained with the primary mutant EGFR polypeptide-specific antibody, washed and labeled with a fluorescent-labeled secondary antibody. The cells would then be analyzed on a flow cytometer (e.g. a Beckman Coulter FC500) according to the specific protocols of the instrument used. Such an analysis would identify the level of expressed mutant EGFR polypeptide in the tumor.

30 Immunohistochemical (IHC) staining may be also employed to determine the expression and/or activation status of mutant EGFR polypeptide in a mammalian cancer (e.g. AML) before, during, and after treatment with a drug targeted at inhibiting EGFR activity.

IHC may be carried out according to well-known techniques. See, e.g., ANTIBODIES; A LABORATORY MANUAL, Chapter 10, Harlow & Lane Eds., Cold Spring Harbor Laboratory (1988). Briefly, and by way of example, paraffin-embedded tissue (e.g. tumor tissue from a biopsy) is prepared for immunohistochemical staining by deparaffinizing tissue sections with xylene followed by ethanol; hydrating in water then PBS; unmasking antigen by heating slide in sodium citrate buffer; incubating sections in hydrogen peroxide; blocking in blocking solution; incubating slide in primary anti-mutant EGFR polypeptide antibody and secondary antibody; and finally detecting using ABC avidin/biotin method according to manufacturer's instructions.

Immunofluorescence (IF) assays may be also employed to determine the expression and/or activation status of mutant EGFR polypeptide in a mammalian cancer before, during, and after treatment with a drug targeted at inhibiting EGFR activity. IF may be carried out according to well-known techniques. See, e.g., J. M. Polak and S. Van Noorden (1997) INTRODUCTION TO IMMUNOCYTOCHEMISTRY, 2nd Ed.; ROYAL MICROSCOPY SOCIETY MICROSCOPY HANDBOOK 37, BioScientific/Springer-Verlag. Briefly, and by way of example, patient samples may be fixed in paraformaldehyde followed by methanol, blocked with a blocking solution such as horse serum, incubated with the primary antibody against mutant EGFR polypeptide followed by a secondary antibody labeled with a fluorescent dye such as Alexa 488 and analyzed with an epifluorescent microscope.

Antibodies employed in the above-described assays may be advantageously conjugated to fluorescent dyes (e.g. Alexa488, PE), or other labels, such as quantum dots, for use in multi-parametric analyses along with other signal transduction (phospho-AKT, phospho-Erk 1/2) and/or cell marker (cytokeratin) antibodies.

A variety of other protocols, including enzyme-linked immunosorbent assay (ELISA), radio-Immunoassay (RIA), and fluorescent-activated cell sorting (FACS), for measuring mutant EGFR polypeptide are known in the art and provide a basis for diagnosing altered or abnormal levels of mutant EGFR polypeptide expression. Normal or standard values for mutant EGFR polypeptide expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to mutant EGFR polypeptide under conditions suitable for complex formation. The amount of standard complex formation may be quantified by various methods, but preferably by photometric means. Quantities of mutant EGFR polypeptide expressed in subject, control, and disease

samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

Similarly, AQUA peptides for the detection/quantification of expressed mutant EGFR polypeptide in a biological sample comprising cells from a tumor may be prepared and used in standard AQUA assays, as described in detail above. Accordingly, in some preferred embodiments of the methods of the invention, the mutant EGFR polypeptide-specific reagent comprises a heavy isotope labeled phosphopeptide (AQUA peptide) corresponding to a peptide sequence comprising the mutant junction of mutant EGFR polypeptide, as described above.

Mutant EGFR polypeptide-specific reagents useful in practicing the methods of the invention may also be mRNA, oligonucleotide or DNA probes that can directly hybridize to, and detect, mutant or truncated polypeptide expression transcripts in a biological sample. Briefly, and by way of example, formalin-fixed, paraffin-embedded patient samples may be probed with a fluorescein-labeled RNA probe followed by washes with formamide, SSC and PBS and analysis with a fluorescent microscope.

Polynucleotides encoding mutant EGFR polypeptide may also be used for diagnostic purposes. The polynucleotides that may be used include oligonucleotide sequences, antisense RNA and DNA molecules. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of mutant EGFR polypeptide or truncated active EGFR polypeptide may be correlated with disease. For example, the diagnostic assay may be used to distinguish between absence, presence, and excess expression of mutant EGFR polypeptide, and to monitor regulation of mutant EGFR polypeptide levels during therapeutic intervention.

In one preferred embodiment, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding mutant EGFR polypeptide or truncated active EGFR polypeptide, or closely related molecules, may be used to identify nucleic acid sequences which encode mutant EGFR polypeptide. The construction and use of such probes is described above. The specificity of the probe, whether it is made from a highly specific region, e.g., 10 unique nucleotides in the mutant junction, or a less specific region, e.g., the 3' coding region, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low) will determine whether the probe

identifies only naturally occurring sequences encoding mutant EGFR polypeptide, alleles, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably contain at least 50% of the nucleotides from any of the mutant EGFR polypeptide encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and derived from the nucleotide sequence and encompassing the G724S, G719S, L858R mutation point, or any combination thereof, or from genomic sequence including promoter, enhancer elements, and introns of the naturally occurring EGFR polypeptides but comprising the G724S, G719S, L858R, or any combination thereof mutation points sequence.

A mutant EGFR polynucleotide or truncated EGFR polynucleotide of the invention may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; or in dip stick, pin, ELISA or chip assays utilizing fluids or tissues from patient biopsies to detect altered EGFR polypeptide expression. Such qualitative or quantitative methods are well known in the art. Mutant EGFR polynucleotides may be labeled by standard methods, and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in the biopsied or extracted sample is significantly altered from that of a comparable control sample, the nucleotide sequences have hybridized with nucleotide sequences in the sample, and the presence of altered levels of nucleotide sequences encoding mutant EGFR polypeptide in the sample indicates the presence of the associated disease. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or in monitoring the treatment of an individual patient.

In order to provide a basis for the diagnosis of disease characterized by expression of mutant EGFR polypeptide, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, which encodes mutant EGFR polypeptide, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with those from an experiment where a known amount of a substantially purified polynucleotide is used. Standard values obtained from normal samples may be compared with values obtained

from samples from patients who are symptomatic for disease. Deviation between standard and subject values is used to establish the presence of disease.

Once disease is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to evaluate whether the level of expression in the patient
5 begins to approximate that which is observed in the normal patient. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

Additional diagnostic uses for mutant EGFR polynucleotides of the invention may involve the use of polymerase chain reaction (PCR), a preferred assay format that is standard
10 to those of skill in the art. See, e.g., MOLECULAR CLONING, A LABORATORY MANUAL, 2nd. edition, Sambrook, J., Fritsch, E. F. and Maniatis, T., eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989). PCR oligomers may be chemically synthesized, generated enzymatically, or produced from a recombinant source. Oligomers will preferably consist of two nucleotide sequences, one with sense orientation (5'
15 to 3') and another with antisense (3' to 5'), employed under optimized conditions for identification of a specific gene or condition. The same two oligomers, nested sets of oligomers, or even a degenerate pool of oligomers may be employed under less stringent conditions for detection and/or quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantitate the expression of mutant EGFR
20 polypeptide include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and standard curves onto which the experimental results are interpolated (Melby *et al.*, J. Immunol. Methods, 159:235-244 (1993); Duplaa *et al.* Anal. Biochem. 229-236 (1993)). The speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and
25 a spectrophotometric or calorimetric response gives rapid quantitation.

Other suitable methods for nucleic acid detection, such as minor groove-binding conjugated oligonucleotide probes (see, e.g. U.S. Pat. No. 6,951,930, "Hybridization-Triggered Fluorescent Detection of Nucleic Acids") are known to those of skill in the art.

Example 1: Identification of a somatic mutation in the P-loop of the tyrosine kinase of EGFR that is predictive of response to EGFR antibody therapy

Fresh frozen adenocarcinoma samples were obtained from resection of a primary colorectal and a carcinoma as well as tissue from the non-affected Vall d'Hebron Hospital in Barcelona Spain. Patients were consented for the trial and tissue collection in a manner consistent with the local ethics board and following all institutional procedures. The tumor sample was subsequently processed at the biological samples platform of the Broad Institute where DNA and RNA are both obtained from the tissue of tumor as well as the noncancerous normal colon. DNA was subjected to standard quality control measures as well as quantification of DNA contents using picogreen.

Following documentation of adequate quality genomic material, the DNA from the tumor sample as well as the matched germline DNA were used for the construction of Illumina sequencing libraries using standard library construction approaches. Following library construction, the tumor and normal each subject to whole genome shotgun sequencing. This sequencing, performed on the Illumina genome analyzer was performed with paired 101-base pair reads with the goal of achieving approximately 30 fold coverage of each the tumor and germline genomes.

The genomic data were processed by the cancer genome analysis group of the Broad Institute. Briefly, the reads from the tumor and normal sequencing were mapped back to the referenced human genome to create BAM files of each genome. These files were subsequently analyzed in the Firehose pipeline which contains distinct algorithms for detection of somatic mutations, somatic insertions and deletions, somatic copy number alterations as well as somatic genomic rearrangements in the tumor sample.

The patient from the clinical trial who demonstrated a partial response to single agent treatment with Cetuximab was subjected to 30x whole genome sequencing of the tumor derived DNA as well as the normal colon. Through review of the genomic data obtained using the Firehose pipeline, 59 missense mutations, five nonsense mutations and 80 high confidence rearrangements were identified. Review of these alterations showed no genomic event that clearly would suggest a rationale for the striking dependence of this tumor upon signaling from EGFR and a rationale for sensitivity to Cetuximab treatment. Given our strong pre-existing hypothesis of genomic rearrangement within EGFR, this locus was manually

reviewed for the presence of a somatic alteration that was missed by the automated calling algorithms.

Upon manual review of EGFR, three mutations in the tyrosine kinase domain were identified- G724S, G719S and L858R. The mutation, G724S, was notably located in a
5 specific subdomain of the tyrosine kinase, the P-loop, which has been noted to be mutated recurrently in non-small cell lung carcinoma. The G724S mutation was present in 10/42 reads from the tumor DNA that covered this specific base. In contrast, in the normal genome a mutation was noted in only one of the genomic reads from sequencing.

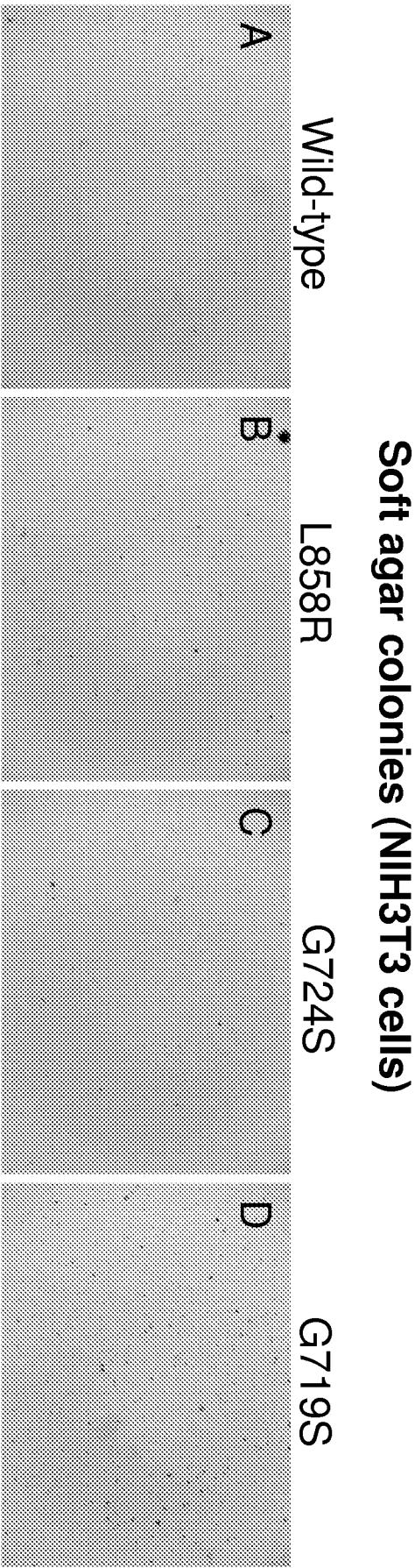
10

OTHER EMBODIMENTS

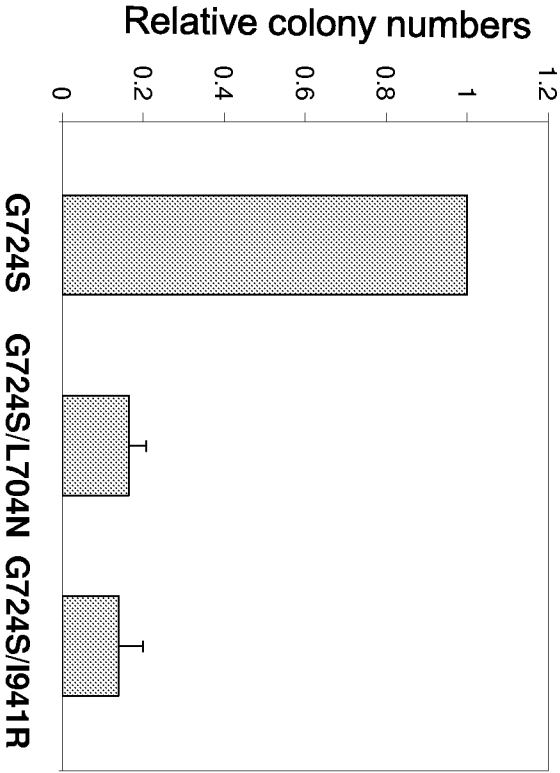
While the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages,
15 and modifications are within the scope of the following claims.

What is claimed is:

1. A method of accessing the effectiveness of to an EGFR antibody therapy treatment regimen of a subject having cancer comprising, obtaining a sample from the subject and detecting the presence or absence of a G724S mutation, G719S mutation , L858R mutation or any combination thereof in the EGFR polypeptide, wherein the presence of one or more of the mutations indicates the subject is responsive to EGFR antibody therapy treatment.
2. The method of claim 1, wherein the EGFR antibody is Cetuximab, Panitumumab, Zalutumumab, Nimotuzumab, Necitumumab and Matuzumab.
3. The method of claim 1, wherein the cancer is colorectal adenocarcinoma or head and neck cancer.
4. The method of claim 1, wherein said subject has not received treatment for the cancer.
5. The method of claim 1, wherein said subject has received treatment for the cancer.
6. The method of claim 1, wherein the method is implemented in a flow-cytometry (FC), immuno-histochemistry (IHC), or immuno-fluorescence (IF) assay format.
7. The method of claim 1, wherein the method is implemented in a polymerase chain reaction (PCR) sequencing assay format.
8. A kit for the detection of mutant EGFR polypeptide in a biological sample, the kit comprising a reagent specifically binds to or detects a mutant G724S EGFR polypeptide, a mutant G719S EGFR polypeptide or a a mutant L858R EGFR polypeptide but does not bind to or detect wild type EGFR and one or more secondary reagents.



E



F

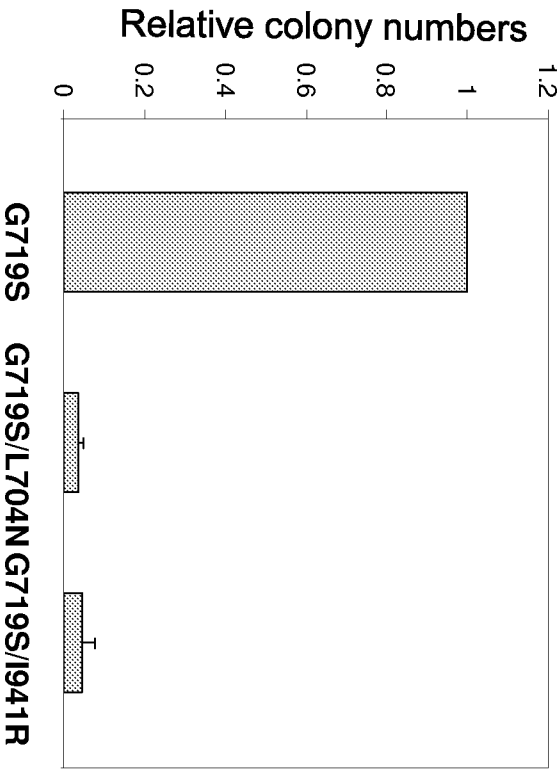


FIG. 1

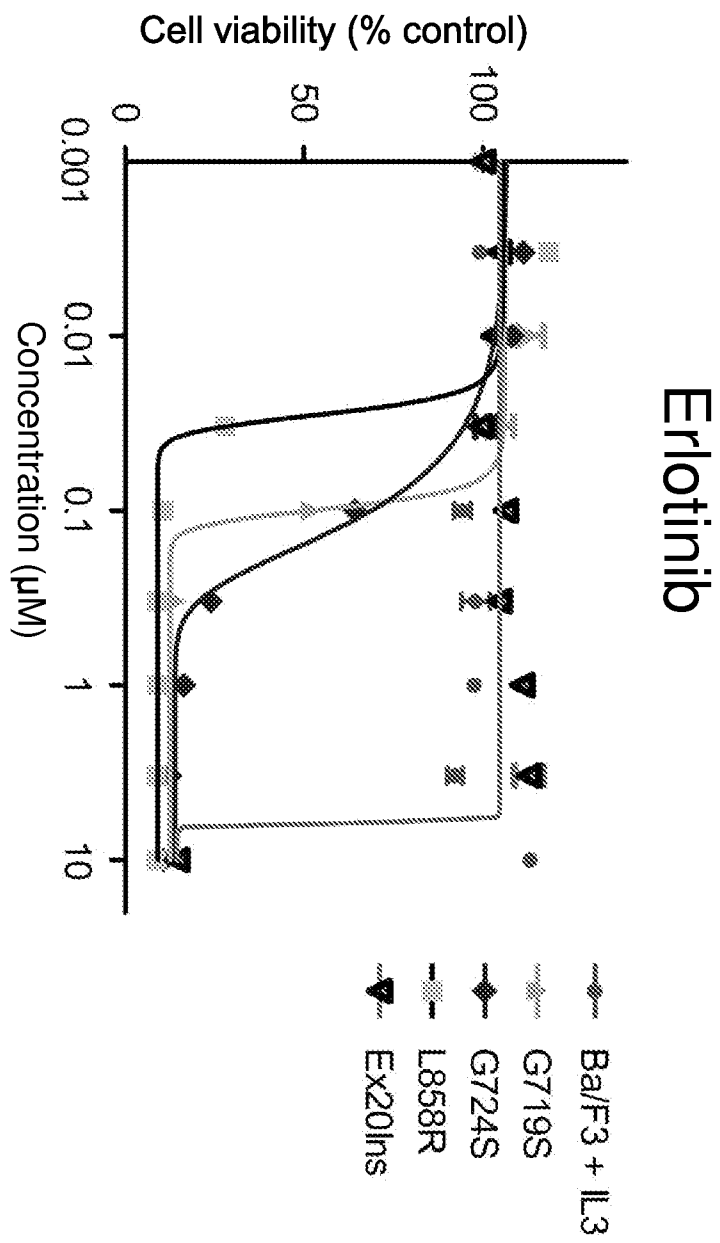


FIG. 2

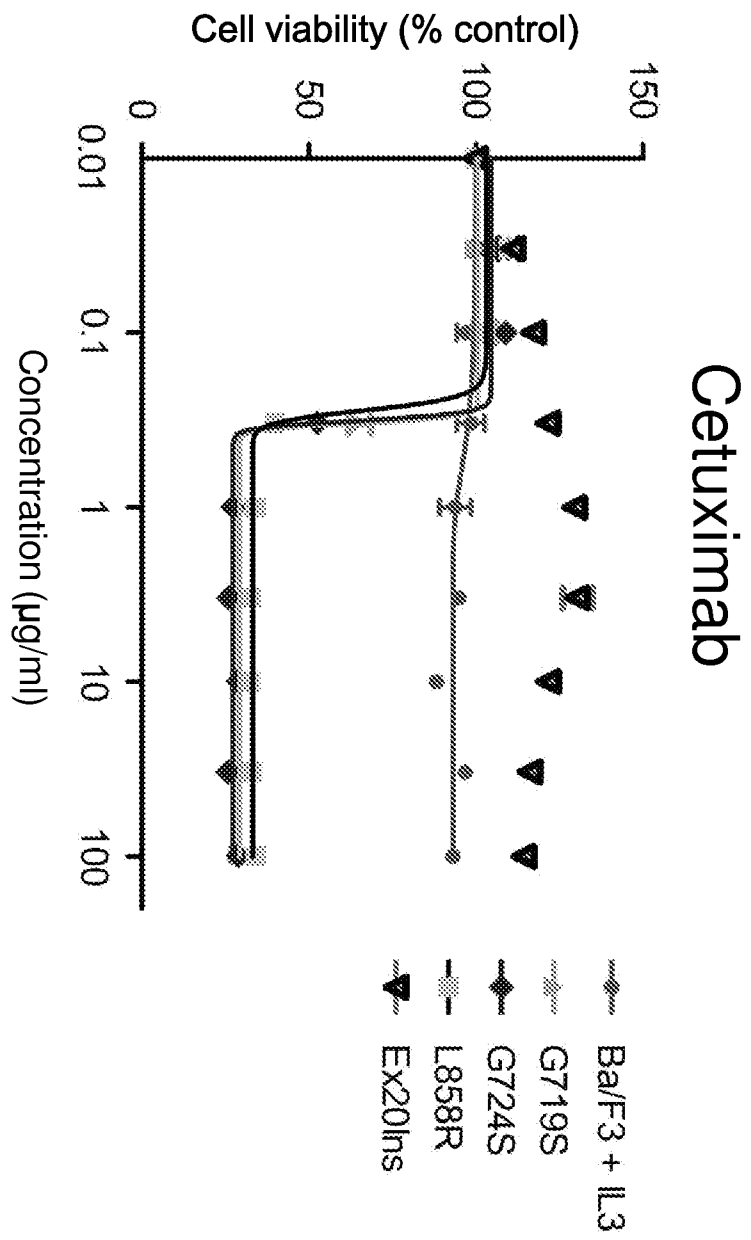


FIG. 3