

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
25 October 2007 (25.10.2007)

PCT

(10) International Publication Number
WO 2007/121017 A2

- (51) International Patent Classification: Not classified
- (21) International Application Number: PCT/US2007/064116
- (22) International Filing Date: 16 March 2007 (16.03.2007)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/784,025 20 March 2006 (20.03.2006) US
- (71) Applicant (for all designated States except AT, US): NO-VARTIS AG [CH/CH]; Lichtstrasse 35, CH-4056 Basel (CH).
- (71) Applicant (for AT only): NOVARTIS PHARMA GMBH [AT/AT]; Brunner Strasse 59, A-1230 Vienna (AT).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): CULVER, Kenneth Wayne [US/US]; 8 Jared Drive, Mendham, New Jersey 07945 (US). ZHU, Jian [CN/US]; 25 Dunnerdale Road, Morris Plains, New Jersey 07950 (US). LILLEBERG, Stan [US/US]; 2121 S. 84th Street, Omaha, Nebraska 68124 (US).
- (74) Agent: PRINCE, John T.; Novartis, Corporate Intellectual Property, One Health Plaza, East Hanover, NJ 07936-1080 (US).
- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:
— without international search report and to be republished upon receipt of that report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: MUTATIONS AND POLYMORPHISMS OF FMS-LIKE TYROSINE KINASE 4

(57) Abstract: This invention relates generally to the analytical testing of tissue samples in vitro, and more particularly to aspects of genetic polymorphisms and mutations of the FLT 4 gene. The invention provides new FLT 4 mutations and SNPs, useful in the diagnosis and treatment of subjects in need thereof. Accordingly, the various aspects of the present invention relate to polynucleotides encoding the FLT 4 mutations of the invention, expression vectors encoding the FLT 4 mutant polypeptides of the invention and organisms that express the FLT 4 mutant and polymorphic polynucleotides and/or FLT 4 mutant/polymorphic polypeptides of the invention. The various aspects of the present invention further relate to diagnostic/theranostic methods and kits that use the FLT 4 mutations and polymorphisms of the invention to identify individuals predisposed to disease or to classify individuals with regard to drug responsiveness, side effects, or optimal drug dose.



WO 2007/121017 A2

-1-

MUTATIONS AND POLYMORPHISMS OF FMS-LIKE TYROSINE KINASE 4

FIELD OF THE INVENTION

[0001] This invention relates generally to the analytical testing of tissue samples *in vitro*, and more particularly to aspects of genetic mutations and polymorphisms of FMS-like tyrosine kinase 4 (FLT-4).

BACKGROUND OF THE INVENTION

[0002] Conventional medical approaches to diagnosis and treatment of disease is based on clinical data alone, or made in conjunction with a diagnostic test. Such traditional practices often lead to therapeutic choices that are not optimal for the efficacy of the prescribed drug therapy or to minimize the likelihood of side effects for an individual subject. Therapy specific diagnostics (*a.k.a.*, theranostics) is an emerging medical technology field, which provides tests useful to diagnose a disease, choose the correct treatment regime and monitor a subject's response. That is, theranostics are useful to predict and assess drug response in individual subjects, *i.e.*, individualized medicine. Theranostic tests are also useful to select subjects for treatments that are particularly likely to benefit from the treatment or to provide an early and objective indication of treatment efficacy in individual subjects, so that the treatment can be altered with a minimum of delay. Theranostics are useful in clinical diagnosis and management of a variety of diseases and disorders, which include, but are not limited to, *e.g.*, cardiovascular disease, cancer, infectious diseases, Alzheimer's disease and the prediction of drug toxicity or drug resistance. Theranostic tests may be developed in any suitable diagnostic testing format, which include, but is not limited to, *e.g.*, immunohistochemical tests, clinical chemistry, immunoassay, cell-based technologies, and nucleic acid tests.

[0003] Progress in pharmacogenomics and pharmacogenetics, which establishes correlations between responses to specific drugs and the genetic profile of individual patients and/or their tumours, is foundational to the development of new theranostic approaches. As such, there is a need in the art for the evaluation of patient-to-patient variations and tumour mutations in gene sequence and gene expression. A common form of genetic profiling relies on the identification of DNA sequence variations called single nucleotide polymorphisms ("SNPs"), which are one type of genetic alteration leading to patient-to-patient variation in individual drug response. In addition, it is well established in the art that acquired DNA

-2-

changes (mutations) are responsible, alone or in part, for pathological processes. It follows that, there is a need art to identify and characterize genetic mutations and SNPs, which are useful to identify the genotypes of subjects and their tumours associated with drug responsiveness, side effects, or optimal dose.

[0004] FMS-like tyrosine kinase 4 (*a.k.a.*, FLT-4; vascular endothelial growth factor receptor 3; VEGFR3) is a receptor-type tyrosine kinase. Galland *et al.*, *Genomics* 13: 475-478, 1992. The deduced amino acid sequence of the intracellular portion of the molecule showed that it was strongly related to FLT1 and KDR and to a lesser degree to members of the class III receptor-type tyrosine kinases: FMS, PDGFR, KIT, and FLT-3. The protein product of the FLT-4 receptor tyrosine kinase cDNA is structurally similar to the FLT-1 and KDR/FLK-1 receptors (Pajusola *et al.*, 1992), but FLT-4 does not bind VEGF (Pajusola *et al.*, *Oncogene* 9: 3545-3555, 1994). Lee *et al.* (1996) identified and characterized a vascular endothelial growth factor-related protein (VEGFC) that specifically binds to the extracellular domain of FLT-4 and stimulates tyrosine phosphorylation and mitogenesis of endothelial cells. Lee *et al.*, *Proc. Nat. Acad. Sci.* 93: 1988-1992, 1996. FLT-4 mutation has been associated with lymphedema (Irrthum *et al.*, *Am. J. Hum. Genet.* 67: 295-301, 2000; Brice *et al.*, *J Med Genet.* 42(2):98-102, 2005) and haemangioma (Walter *et al.*, *Genes Chromosomes Cancer* 33: 295-303, 2002). Accordingly, there is a need in the art for additional information about the relationship between FLT-4 mutations and cancer.

SUMMARY OF THE INVENTION

[0005] The invention provides for the use of an FLT-4 modulating agent in the manufacture of a medicament for the treatment of cancer in a selected patient population. The patient population is selected on the basis of the genotype of the patients at an FLT-4 genetic locus indicative of efficacy of the FLT-4 modulating agent in treating cancer. In several embodiments, the cancer can be colon cancer.

[0006] The invention also provides an isolated polynucleotide having a sequence encoding an FLT-4 mutation. In several embodiments, the FLT-4 mutations are the previously-identified mutations listed in TABLE 1. Accordingly, the invention provides vectors and organisms containing the FLT-4 mutations of the invention and polypeptides encoded by polynucleotides containing the FLT-4 mutations of the invention.

-3-

[0007] The invention further provides a method for treating cancer in a subject. The genotype or haplotype of a subject is obtained at an FLT-4 gene locus, so that the genotype and/or haplotype is indicative of a propensity of the cancer to respond to the drug. Then, an anti-cancer therapy is administered to the subject.

[0008] The invention provides a method for diagnosing cancer in a subject and a method for choosing subjects for inclusion in a clinical trial for determining efficacy of an FLT-4 modulating agent; in both these methods the genotype and/or haplotype of a subject is interrogated at an FLT-4 gene locus. Also provided by the invention are kits for use in determining a treatment strategy for cancer.

[0009] The invention also provides for the use of each of the mutations of the inventions as a drug target.

DETAILED DESCRIPTION OF THE INVENTION

[0010] It is to be appreciated that certain aspects, modes, embodiments, variations and features of the invention are described below in various levels of detail in order to provide a substantial understanding of the present invention. In general, such disclosure provides new FLT-4 mutations and SNPs that may be useful, alone or in combination, in the diagnosis and treatment of subjects in need thereof. Accordingly, the various aspects of the present invention relate to polynucleotides encoding FLT-4 mutations and polymorphisms of the invention, expression vectors encoding the FLT-4 mutant polypeptides of the invention and organisms that express the FLT-4 mutant/polymorphic polynucleotides and/or FLT-4 mutant/polymorphic polypeptides of the invention. The various aspects of the present invention further relate to diagnostic/theranostic methods and kits that use the FLT-4 mutations and/or polymorphisms of the invention to identify individuals predisposed to disease or to classify individuals and tumours with regard to drug responsiveness, side effects, or optimal drug dose. In other aspects, the invention provides methods for compound validation and a computer system for storing and analyzing data related to the FLT-4 mutations and polymorphisms of the invention. Accordingly, various particular embodiments that illustrate these aspects follow.

[0011] *Definitions.* The definitions of certain terms as used in this specification are provided below. Definitions of other terms may be found in the glossary provided by the U.S. Department of Energy, Office of Science, Human Genome Project (http://www.ornl.gov/sci/techresources/Human_Genome/glossary/).

-4-

[0012] As used herein, the term "allele" means a particular form of a gene or DNA sequence at a specific chromosomal location (locus).

[0013] As used herein, the term "antibody" includes, but is not limited to, polyclonal antibodies, monoclonal antibodies, humanized or chimaeric antibodies and biologically functional antibody fragments sufficient for binding of the antibody fragment to the protein.

[0014] As used herein, the term "clinical response" means any or all of the following: a quantitative measure of the response, no response, and adverse response (*i.e.*, side effects).

[0015] As used herein, the term "clinical trial" means any research study designed to collect clinical data on responses to a particular treatment, and includes but is not limited to phase I, phase II and phase III clinical trials. Standard methods are used to define the patient population and to enrol subjects.

[0016] As used herein, the term "effective amount" of a compound is a quantity sufficient to achieve a desired pharmacodynamic, toxicologic, therapeutic and/or prophylactic effect, for example, an amount which results in the prevention of or a decrease in the symptoms associated with a disease that is being treated, *e.g.*, the diseases associated with FLT-4 mutant polypeptides and FLT-4 mutant polynucleotides identified herein. The amount of compound administered to the subject will depend on the type and severity of the disease and on the characteristics of the individual, such as general health, age, sex, body weight and tolerance to drugs. It will also depend on the degree, severity and type of disease. The skilled artisan will be able to determine appropriate dosages depending on these and other factors. Typically, an effective amount of the compounds of the present invention, sufficient for achieving a therapeutic or prophylactic effect, range from about 0.000001 mg per kilogram body weight per day to about 10,000 mg per kilogram body weight per day. Preferably, the dosage ranges are from about 0.0001 mg per kilogram body weight per day to about 100 mg per kilogram body weight per day. The compounds of the present invention can also be administered in combination with each other, or with one or more additional therapeutic compounds.

[0017] Glivec® (Gleevec®; imatinib) is a medication for chronic myeloid leukaemia (CML) and certain stages of gastrointestinal stromal tumours (GIST). It targets and interferes with the molecular abnormalities that drive the growth of cancer cells. Corless CL *et al.*, *J. Clin. Oncol.* 22(18):3813-25 (September 15, 2004); Verweij J *et al.*, *Lancet* 364(9440):1127-34 (September 25, 2004); Kantarjian HM *et al.*, *Blood* 104(7):1979-88 (October 1, 2004). By

inhibiting multiple targets, Glivec® has potential as an anticancer therapy for several types of cancer, including leukaemia and solid tumours.

[0018] The aromatase inhibitor FEMARA® is a treatment for advanced breast cancer in postmenopausal women. It blocks the use of oestrogen by certain types of breast cancer that require oestrogen to grow. Janicke F, *Breast* 13 Suppl 1:S10-8 (December 2004); Mouridsen H *et al.*, *Oncologist* 9(5):489-96 (2004).

[0019] Sandostatin® LAR® is used to treat patients with acromegaly and to control symptoms, such as severe diarrhoea and flushing, in patients with functional gastro-entero-pancreatic (GEP) tumours (*e.g.*, metastatic carcinoid tumours and vasoactive intestinal peptide-secreting tumours [VIPomas]). Oberg K, *Chemotherapy* 47 Suppl 2:40-53 (2001); Raderer M *et al.*, *Oncology* 60(2):141-5 (2001); Aparicio T *et al.*, *Eur. J. Cancer* 37(8):1014-9 (May 2001). Sandostatin® LAR® regulates hormones in the body to help manage diseases and their symptoms.

[0020] ZOMETA® is a treatment for hypocalcaemia of malignancy (HCM)1 and for the treatment of bone metastases across a broad range of tumour types. These tumours include multiple myeloma, prostate cancer, breast cancer, lung cancer, renal cancer and other solid tumours. Rosen LS *et al.*, *Cancer* 100(12):2613-21 (June 15, 2004).

[0021] Vatalanib (1-[4-chloroanilino]-4-[4-pyridylmethyl] phthalazine succinate) is a multi-VEGF receptor (VEGF) inhibitor that may block the creation of new blood vessels to prevent tumour growth. This compound inhibits all known VEGF receptor tyrosine kinases, blocking angiogenesis and lymphangiogenesis. Dreys J *et al.*, *Cancer Res.* 60:4819-4824 (2000); Wood JM *et al.*, *Cancer Res.* 60:2178-2189 (2000). Vatalanib is being studied in two large, multinational, randomized, phase III, placebo-controlled trials in combination with FOLFOX-4 in first-line and second-line treatment of patients with metastatic colorectal cancer. Thomas A *et al.*, *37th Annual Meeting of the American Society of Clinical Oncology*, San Francisco, CA, Abstract 279 (May 12-15, 2001).

[0022] The orally bioavailable rapamycin derivative everolimus inhibits oncogenic signalling in tumour cells. By blocking the mammalian target of rapamycin (mTOR)-mediated signalling, everolimus exhibits broad antiproliferative activity in tumour cell lines and animal models of cancer. Boulay A *et al.*, *Cancer Res.* 64:252-261 (2004). In preclinical studies, everolimus also potently inhibited the proliferation of human umbilical vein endothelial cells directly indicating an involvement in angiogenesis. By blocking tumour cell

-6-

proliferation and angiogenesis, everolimus may provide a clinical benefit to patients with cancer. Everolimus is being investigated for its antitumour properties in a number of clinical studies in patients with haematological and solid tumours. Huang S & Houghton PJ, *Curr. Opin. Investig. Drugs* 3:295-304 (2002).

[0023] Gimatecan is a novel oral inhibitor of topoisomerase I (topo I). Gimatecan blocks cell division in cells that divide rapidly, such as cancer cells, which activates apoptosis. Preclinical data indicate that gimatecan is not a substrate for multidrug resistance pumps, and that it increases the drug-target interaction. De Cesare M *et al.*, *Cancer Res.* 61:7189-7195 (2001). Phase I clinical studies indicate that the dose-limiting toxicity of gimatecan is myelosuppression.

[0024] Patupilone is a microtubule stabilizer. Altmann K-H, *Curr. Opin. Chem. Biol.* 5:424-431 (2001); Altmann K-H *et al.*, *Biochim. Biophys. Acta* 470:M79-M91 (2000); O'Neill V *et al.*, *36th Annual Meeting of the American Society of Clinical Oncology*; May 19-23, 2000; New Orleans, LA, Abstract 829; Calvert PM *et al.* *Proceedings of the 11th National Cancer Institute-European Organization for Research and Treatment of Cancer/American Association for Cancer Research Symposium on New Drugs in Cancer Therapy*; November 7-10, 2000; Amsterdam, The Netherlands, Abstract 575. Patupilone blocked mitosis and induced apoptosis greater than the frequently used anticancer drug paclitaxel. Also, patupilone retained full activity against human cancer cells that were resistant to paclitaxel and other chemotherapeutic agents.

[0025] Midostaurin is an inhibitor of multiple signalling proteins. By targeting specific receptor tyrosine kinases and components of several signal transduction pathways, midostaurin impacts several targets involved in cell growth (*e.g.*, KIT, PDGFR, PKC), leukaemic cell proliferation (*e.g.*, FLT3), and angiogenesis (*e.g.*, VEGFR2). Weisberg E *et al.* *Cancer Cell* 1:433-443 (2002); Fabbro D *et al.*, *Anticancer Drug Des.* 15:17-28 (2000). In preclinical studies, midostaurin showed broad antiproliferative activity against various tumour cell lines, including those that were resistant to several other chemotherapeutic agents.

[0026] The somatostatin analogue pasireotide is a stable cyclohexapeptide with broad somatotropin release inhibiting factor (SRIF) receptor binding. Bruns C *et al.*, *Eur. J. Endocrinol.* 146(5):707-16 (May 2002); Weckbecker G *et al.*, *Endocrinology* 143(10):4123-30 (October 2002); Oberg K, *Chemotherapy* 47 Suppl 2:40-53 (2001).

-7-

[0027] LBH589 is a histone deacetylase (HDAC) inhibitor. By blocking the deacetylase activity of HDAC, HDAC inhibitors activate gene transcription of critical genes that cause apoptosis (programmed cell death). By triggering apoptosis, LBH589 induces growth inhibition and regression in tumour cell lines. LBH589 is being tested in phase I clinical trials as an anticancer agent. See also, George P *et al.*, *Blood* 105(4):1768-76 (February 15, 2005).

[0028] AEE788 inhibits multiple receptor tyrosine kinases including EGFR, HER2, and VEGFR, which stimulate tumour cell growth and angiogenesis. Traxler P *et al.*, *Cancer Res.* 64:4931-4941 (2004). In preclinical studies, AEE788 showed high target specificity and demonstrated antiproliferative effects against tumour cell lines and in animal models of cancer. AEE788 also exhibited direct antiangiogenic activity. AEE788 is currently in phase I clinical development.

[0029] AMN107 is an oral tyrosine kinase inhibitor that targets BCR-ABL, KIT, and PDGFR. Preclinical studies have shown in cellular assays using Philadelphia chromosome-positive (Ph+) CML cells that AMN107 is highly potent and has high selectivity for BCR-ABL, KIT, and PDGFR. Weisberg E *et al.*, *Cancer Cell* 7(2):129-41 (February 2005); O'Hare T *et al.*, *Cancer Cell* 7(2):117-9 (February 2005). AMN107 also shows activity against mutated variants of BCR-ABL. AMN107 is currently being studied in phase I clinical trials.

[0030] As used herein, the term "FLT-4 modulating agent" is any compound that alters (*e.g.*, increases or decreases). The expression level or biological activity level of FLT-4 polypeptide compared to the expression level or biological activity level of FLT-4 polypeptide in the absence of the FLT-4 modulating agent. FLT-4 modulating agent can be a small molecule, antibody, polypeptide, carbohydrate, lipid, nucleotide, or combination thereof. The FLT-4 modulating agent can be an organic compound or an inorganic compound.

[0031] As used herein, "expression" includes but is not limited to one or more of the following: transcription of the gene into precursor mRNA; splicing and other processing of the precursor mRNA to produce mature mRNA; mRNA stability; translation of the mature mRNA into protein (including codon usage and tRNA availability); and glycosylation and/or other modifications of the translation product, if required for proper expression and function.

-8-

[0032] As used herein, the term “gene” means a segment of DNA that contains all the information for the regulated biosynthesis of an RNA product, including promoters, exons, introns, and other untranslated regions that control expression.

[0033] As used herein, the term “genotype” means an unphased 5’ to 3’ sequence of nucleotide pairs found at one or more polymorphic or mutant sites in a locus on a pair of homologous chromosomes in an individual. As used herein, genotype includes a full-genotype and/or a sub-genotype.

[0034] As used herein, the term “locus” means a location on a chromosome or DNA molecule corresponding to a gene or a physical or phenotypic feature.

[0035] As used herein, the term “mutant” means any heritable or acquired variation from the wild-type that alters the nucleotide sequence thereby changing the protein sequence. The term “mutant” is used interchangeably with the terms “marker”, “biomarker”, and “target” throughout the specification.

[0036] As used herein, the term “medical condition” includes, but is not limited to, any condition or disease manifested as one or more physical and/or psychological symptoms for which treatment and/or prevention is desirable, and includes previously and newly identified diseases and other disorders.

[0037] As used herein, the term “nucleotide pair” means the two nucleotides bound to each other between the two nucleotide strands.

[0038] As used herein, the term “polymorphic site” means a position within a locus at which at least two alternative sequences are found in a population, the most frequent of which has a frequency of no more than 99%.

[0039] As used herein, the term “polymorphism” means any sequence variant present at a frequency of >1% in a population. The sequence variant may be present at a frequency significantly greater than 1% such as 5% or 10% or more. Also, the term may be used to refer to the sequence variation observed in an individual at a polymorphic site. Polymorphisms include nucleotide substitutions, insertions, deletions and microsatellites and may, but need not, result in detectable differences in gene expression or protein function.

[0040] As used herein, the term “polynucleotide” means any RNA or DNA, which may be unmodified or modified RNA or DNA. Polynucleotides include, without limitation, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, RNA that is mixture of single- and double-stranded

regions, and hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, polynucleotide refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The term polynucleotide also includes DNAs or RNAs containing one or more modified bases and DNAs or RNAs with backbones modified for stability or for other reasons. In a particular embodiment, the polynucleotide contains polynucleotide sequences from the FLT-4 gene.

[0041] As used herein, the term "polypeptide" means any polypeptide comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, *i.e.*, peptide isosteres. Polypeptide refers to both short chains, commonly referred to as peptides, glycopeptides or oligomers, and to longer chains, generally referred to as proteins.

Polypeptides may contain amino acids other than the 20 gene-encoded amino acids.

Polypeptides include amino acid sequences modified either by natural processes, such as post-translational processing, or by chemical modification techniques that are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. In a particular embodiment, the polypeptide contains polypeptide sequences from the FLT-4 protein.

[0042] As used herein, the term "small molecule" means a composition that has a molecular weight of less than about 5 kDa and more preferably less than about 2 kDa. Small molecules can be, *e.g.*, nucleic acids, peptides, polypeptides, glycopeptides, peptidomimetics, carbohydrates, lipids, lipopolysaccharides, combinations of these, or other organic or inorganic molecules.

[0043] As used herein, the term "mutant nucleic acid" means a nucleic acid sequence, which comprises a nucleotide that is variable within an otherwise identical nucleotide sequence between individuals or groups of individuals, thus, existing as alleles. Such mutant nucleic acids are preferably from about 15 to about 500 nucleotides in length. The mutant nucleic acids may be part of a chromosome, or they may be an exact copy of a part of a chromosome, *e.g.*, by amplification of such a part of a chromosome through PCR or through cloning. The mutant probes according to the invention are oligonucleotides that are complementary to a mutant nucleic acid.

[0044] As used herein, the term "SNP nucleic acid" means a nucleic acid sequence, which comprises a nucleotide that is variable within an otherwise identical nucleotide sequence

-10-

between individuals or groups of individuals, thus, existing as alleles. Such SNP nucleic acids are preferably from about 15 to about 500 nucleotides in length. The SNP nucleic acids may be part of a chromosome, or they may be an exact copy of a part of a chromosome, *e.g.*, by amplification of such a part of a chromosome through PCR or through cloning. The SNP nucleic acids are referred to hereafter simply as “SNPs”. The SNP probes according to the invention are oligonucleotides that are complementary to a SNP nucleic acid. In a particular embodiment, the SNP is in the FLT-4 gene.

[0045] As used herein, the term “subject” means that preferably the subject is a mammal, such as a human, but can also be an animal, *e.g.*, domestic animals (*e.g.*, dogs, cats and the like), farm animals (*e.g.*, cows, sheep, pigs, horses and the like) and laboratory animals (*e.g.*, monkey (*e.g.*, cynomolgous monkey), rats, mice, guinea pigs and the like).

[0046] As used herein, the administration of an agent or drug to a subject or patient includes self-administration and the administration by another. It is also to be appreciated that the various modes of treatment or prevention of medical conditions as described are intended to mean “substantial”, which includes total but also less than total treatment or prevention, and wherein some biologically or medically relevant result is achieved.

[0047] The details of one or more embodiments of the invention are set forth in the accompanying description below. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are now described. Other features, objects, and advantages of the invention will be apparent from the description and the claims. In the specification and the appended claims, the singular forms include plural referents unless the context clearly dictates otherwise. Unless defined otherwise, 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. All references cited herein are incorporated herein by reference in their entireties and for all purposes to the same extent as if each individual publication, patent, or patent application was specifically and individually incorporated by reference in its entirety for all purposes.

-11-

[0048] *FLT-4 Mutations and Polymorphisms of the Invention.* To investigate FLT-4 mutations in association with cancer DHPLC (Lilleberg S.L., *Curr. Opin. Drug. Discov. Devel.*, 6(2):237-52 (2003)) analysis was conducted on blood samples from colon cancer patients. Five (5) missense mutations were identified as summarized below in TABLE 1 and TABLE 2 below. Computational analyses were designed to evaluate the effect of these mutations on FLT-4 function.

TABLE 1
FLT-4 Mutations Identified in Colon Cancer Patients

<u>Exon</u>	<u>Mutation/SNP</u>	<u>Allelic Frequency</u>	<u>Unmutated Sequence</u>	<u>Mutated Sequence</u>
Exon 4	AAC>GAC N149D	Hetero- zygous	CTCTTGGTCAACAGGA AGG SEQ ID NO:1	CTCTTGGTCGACAGGA AGG SEQ ID NO:2
Exon 21	ATG>GTG M963V	0.55	CTTCCGCGCCATGGTGG AGCTCG SEQ ID NO:3	CTTCCGCGCCGTGGTGG AGCTCG SEQ ID NO:4
Exon 21	GAT>GAG D970E	0.03	CAGGCTGGATCGGAGG CGGC SEQ ID NO:5	CAGGCTGGAGCGGAGG CGGC SEQ ID NO:6
Exon 25	Ins 60bp E1123-A1143	0.25	60 bp insertion	60 bp insertion
Exon 25	CCG>TCG P1114S	0.2	GGGCCTCCCCGTACCCT GGG SEQ ID NO:7	GGGCCTCCTCGTACCCT GGG SEQ ID NO:8

[0050] As shown above in TABLE 1 and further summarized below in TABLE 2, five (5) missense mutations were identified were identified in the present invention.

TABLE 2
FLT-4 Mutations in Cancer Patients

<u>Gene</u>	<u>Cancer</u>	<u>NT change</u>	<u>Mutation/SNP</u>	<u>Alle. Freq.</u>	<u>Obs.</u>
FLT-4	Colon	AAC>GAC	N149D	het	2
FLT-4	Colon	ATG>GTG	M963V	0.55	1
FLT-4	Colon	GAT>GAG	D970E	0.03	1
FLT-4	Colon	Ins 60bp	E1123-A1143	0.25	1
FLT-4	Colon	CCG>TCG	P1114S	0.2	1

[0051] The missense mutations were located in the human FLT-4 gene (NT 077451). Bioinformatics analysis of the FLT-4 mutations of the invention are further detailed in EXAMPLE I.

[0052] *Identification of FLT-4 Mutations and Polymorphisms of the Invention in Human Cancers. Identification and Characterization of Gene Sequence Variation.* Sequence variation in the human germline consists primarily of SNPs, the remainder being short tandem

-12-

repeats (including micro-satellites), long tandem repeats (mini-satellites), and other insertions and deletions. A SNP is the occurrence of nucleotide variability at a single position in the genome, in which two alternative bases occur at appreciable frequency (*i.e.*, >1%) in the human population. A SNP may occur within a gene or within intergenic regions of the genome.

[0054] Due to their prevalence and widespread nature, SNPs have the potential to be important tools for locating genes that are involved in human disease conditions. *See e.g.*, Wang *et al.*, *Science* 280: 1077-1082 (1998)).

[0055] An association between SNP's and/or mutations and a particular phenotype (*e.g.*, cancer type) does not necessarily indicate or require that the SNP or mutation is causative of the phenotype. Instead, an association with a SNP may merely be due to genome proximity between a SNP and those genetic factors actually responsible for a given phenotype, such that the SNP and said genetic factors are closely linked. That is, a SNP may be in linkage disequilibrium ("LD") with the "true" functional variant. LD exists when alleles at two distinct locations of the genome are more highly associated than expected. Thus, a SNP may serve as a marker that has value by virtue of its proximity to a mutation or other DNA alteration (*e.g.*, gene duplication) that causes a particular phenotype.

[0056] SNPs and mutations that are associated with disorders may also have a direct effect on the function of the genes in which they are located. For example, a sequence variant (*e.g.*, SNP) may result in an amino acid change or may alter exon-intron splicing, thereby directly modifying the relevant protein, or it may exist in a regulatory region, altering the cycle of expression or the stability of the mRNA (*see, e.g.*, Nowotny *et al.*, *Current Opinions in Neurobiology*, 11:637-641 (2001)).

[0057] In describing the polymorphic and mutant sites of the invention, reference is made to the sense strand of the gene for convenience. As recognized by the skilled artisan, however, nucleic acid molecules containing the gene may be complementary double stranded molecules and thus reference to a particular site on the sense strand refers as well to the corresponding site on the complementary antisense strand. That is, reference may be made to the same polymorphic or mutant site on either strand and an oligonucleotide may be designed to hybridize specifically to either strand at a target region containing the polymorphic and/or mutant site. Thus, the invention also includes single-stranded polynucleotides and mutations that are complementary to the sense strand of the genomic variants described herein.

[0058] *Identification and Characterization of SNPs and Mutations.* Many different techniques can be used to identify and characterize SNPs and mutations, including single-strand conformation polymorphism (SSCP) analysis, heteroduplex analysis by denaturing high-performance liquid chromatography (DHPLC), direct DNA sequencing and computational methods (Shi *et al.*, *Clin. Chem.* 47:164-172 (2001)). There is a wealth of sequence information in public databases; computational tools useful to identify SNPs *in silico* by aligning independently submitted sequences for a given gene (either cDNA or genomic sequences). The most common SNP-typing methods currently include hybridization, primer extension, and cleavage methods. Each of these methods must be connected to an appropriate detection system. Detection technologies include fluorescent polarization (Chan *et al.*, *Genome Res.* 9:492-499 (1999)), luminometric detection of pyrophosphate release (pyrosequencing) (Ahmadiian *et al.*, *Anal. Biochem.* 280:103-10 (2000)), fluorescence resonance energy transfer (FRET)-based cleavage assays, DHPLC, and mass spectrometry (Shi, *Clin Chem* 47:164-172 (2001); U.S. Pat. No. 6,300,076 B1). Other methods of detecting and characterizing SNPs and mutations are those disclosed in U.S. Pat. Nos. 6,297,018 B1 and 6,300,063 B1.

[0059] In a particularly preferred embodiment, the detection of polymorphisms and mutations is detected using INVADER™ technology (available from Third Wave Technologies Inc. Madison, Wisconsin USA). In this assay, a specific upstream “invader” oligonucleotide and a partially overlapping downstream probe together form a specific structure when bound to complementary DNA template. This structure is recognized and cut at a specific site by the Cleavase enzyme, resulting in the release of the 5' flap of the probe oligonucleotide. This fragment then serves as the “invader” oligonucleotide with respect to synthetic secondary targets and secondary fluorescently labelled signal probes contained in the reaction mixture. This results in specific cleavage of the secondary signal probes by the Cleavase enzyme. Fluorescent signal is generated when this secondary probe (labelled with dye molecules capable of fluorescence resonance energy transfer) is cleaved. Cleavases have stringent requirements relative to the structure formed by the overlapping DNA sequences or flaps and can, therefore, be used to specifically detect single base pair mismatches immediately upstream of the cleavage site on the downstream DNA strand. Ryan D *et al.*, *Molecular Diagnosis* 4(2): 135-144 (1999) and Lyamichev V *et al.* *Nature Biotechnology* 17: 292-296 (1999), see also U.S. Pat. Nos. 5,846,717 and 6,001,567.

-14-

[0060] The identity of polymorphisms and mutations may also be determined using a mismatch detection technique including, but not limited to, the RNase protection method using riboprobes (Winter *et al.*, *Proc. Natl. Acad. Sci. USA* 82:7575 (1985); Meyers *et al.*, *Science* 230:1242 (1985)) and proteins which recognize nucleotide mismatches, such as the *E. coli* mutS protein (Modrich P, *Ann Rev Genet* 25:229-253 (1991)). Alternatively, variant alleles can be identified by single strand conformation polymorphism (SSCP) analysis (Orita *et al.*, *Genomics* 5:874-879 (1989); Humphries *et al.*, in *Molecular Diagnosis of Genetic Diseases*, Elles R, ed. (1996) pp. 321-340) or denaturing gradient gel electrophoresis (DGGE) (Wartell *et al.*, *Nucl. Acids Res.* 18:2699-2706 (1990); Sheffield *et al.*, *Proc. Natl. Acad. Sci. USA* 86: 232-236 (1989)). A polymerase-mediated primer extension method may also be used to identify the polymorphisms/mutations. Several such methods have been described in the patent and scientific literature and include the "Genetic Bit Analysis" method (WO 92/15712) and the ligase/polymerase mediated genetic bit analysis (U.S. Pat. No. 5,679,524). Related methods are disclosed in WO 91/02087, WO 90/09455, WO 95/17676, and U.S. Pat. Nos. 5,302,509 and 5,945,283. Extended primers containing a polymorphism or mutation may be detected by mass spectrometry as described in U.S. Pat. No. 5,605,798. Another primer extension method is allele-specific PCR. Ruafo *et al.*, *Nucl. Acids Res.* 17: 8392 (1989); Ruafo *et al.*, *Nucl. Acids Res.* 19: 6877-6882 (1991); WO 93/22456; Turki *et al.*, *J. Clin. Invest.* 95: 1635-1641 (1995). In addition, multiple polymorphic and/or mutant sites may be investigated by simultaneously amplifying multiple regions of the nucleic acid using sets of allele-specific primers as described in WO 89/10414.

[0061] *Haplotyping and Genotyping Oligonucleotides.* The invention provides methods and compositions for haplotyping and/or genotyping the genetic polymorphisms (and possibly mutations) in an individual. As used herein, the terms "genotype" and "haplotype" mean the genotype or haplotype containing the nucleotide pair or nucleotide, respectively, that is present at one or more of the novel polymorphic (or mutant) sites described herein and may optionally also include the nucleotide pair or nucleotide present at one or more additional polymorphic (or mutant) sites in the gene. The additional polymorphic (and mutant) sites may be currently known polymorphic/mutant sites or sites that are subsequently discovered.

[0062] The compositions contain oligonucleotide probes and primers designed to specifically hybridize to one or more target regions containing, or that are adjacent to, a polymorphic or mutant site. Oligonucleotide compositions of the invention are useful in

methods for genotyping and/or haplotyping a gene in an individual. The methods and compositions for establishing the genotype or haplotype of an individual at the novel polymorphic/mutant sites described herein are useful for studying the effect of the polymorphisms and mutations in the aetiology of diseases affected by the expression and function of the protein, studying the efficacy of drugs targeting, predicting individual susceptibility to diseases affected by the expression and function of the protein and predicting individual responsiveness to drugs targeting the gene product.

[0063] Some embodiments of the invention contain two or more differently labelled genotyping oligonucleotides, for simultaneously probing the identity of nucleotides at two or more polymorphic or mutant sites. It is also contemplated that primer compositions may contain two or more sets of allele-specific primer pairs to allow simultaneous targeting and amplification of two or more regions containing a polymorphic or mutant site.

[0064] Genotyping oligonucleotides of the invention may be immobilized on or synthesized on a solid surface such as a microchip, bead, or glass slide (see, *e.g.*, WO 98/20020 and WO 98/20019). Such immobilized genotyping oligonucleotides may be used in a variety of polymorphism and mutation detection assays, including but not limited to probe hybridization and polymerase extension assays. Immobilized genotyping oligonucleotides of the invention may comprise an ordered array of oligonucleotides designed to rapidly screen a DNA sample for polymorphisms and mutations in multiple genes at the same time.

[0065] An allele-specific oligonucleotide primer of the invention has a 3' terminal nucleotide, or preferably a 3' penultimate nucleotide, that is complementary to only one nucleotide of a particular SNP, thereby acting as a primer for polymerase-mediated extension only if the allele containing that nucleotide is present. Allele-specific oligonucleotide (ASO) primers hybridizing to either the coding or noncoding strand are contemplated by the invention. An ASO primer for detecting gene polymorphisms and mutations can be developed using techniques known to those of skill in the art.

[0066] Other genotyping oligonucleotides of the invention hybridize to a target region located one to several nucleotides downstream of one of the novel polymorphic or mutant sites identified herein. Such oligonucleotides are useful in polymerase-mediated primer extension methods for detecting one of the novel polymorphisms or mutations described herein and therefore such genotyping oligonucleotides are referred to herein as "primer-extension oligonucleotides". In a preferred embodiment, the 3'-terminus of a primer-

-16-

extension oligonucleotide is a deoxynucleotide complementary to the nucleotide located immediately adjacent to the polymorphic/mutant site.

[0067] *Direct Genotyping Method of the Invention.* One embodiment of a genotyping method of the invention involves isolating from an individual a nucleic acid mixture comprising at least one copy of the gene of interest and/or a fragment or flanking regions thereof, and determining the identity of the nucleotide pair at one or more of the polymorphic/mutant sites in the nucleic acid mixture. As will be readily understood by the skilled artisan, the two "copies" of a germline gene in an individual may be the same on each allele or may be different on each allele. In a particularly preferred embodiment, the genotyping method comprises determining the identity of the nucleotide pair at each polymorphic and mutant site.

[0068] Typically, the nucleic acid mixture is isolated from a biological sample taken from the individual, such as a blood sample, tumour or tissue sample. Suitable tissue samples include whole blood, tumour or as part of any tissue type, semen, saliva, tears, urine, faecal material, sweat, buccal smears, skin and hair. The nucleic acid mixture may be comprised of genomic DNA, mRNA, or cDNA and, in the latter two cases, the biological sample must be obtained from an organ in which the gene may be expressed. Furthermore, it will be understood by the skilled artisan that mRNA or cDNA preparations would not be used to detect polymorphisms or mutations located in introns or in 5' and 3' nontranscribed regions. If a gene fragment is isolated, it must usually contain the polymorphic and/or mutant sites to be genotyped. Exceptions can include mutations leading to truncation of the gene where a specific polymorphism may be lost. In these cases, the specific DNA alterations are determined by assessing the flanking sequences of the gene and underscore the need to specifically look for both polymorphisms and mutations.

[0069] *Direct Haplotyping Method of the Invention.* One embodiment of the haplotyping method of the invention comprises isolating from an individual a nucleic acid molecule containing only one of the two copies of a gene of interest, or a fragment thereof, and determining the identity of the nucleotide at one or more of the polymorphic or mutant sites in that copy. The nucleic acid may be isolated using any method capable of separating the two copies of the gene or fragment. As will be readily appreciated by those skilled in the art, any individual clone will only provide haplotype information on one of the two gene copies present in an individual. If haplotype information is desired for the individual's other copy,

-17-

additional clones will need to be examined. Typically, at least five clones should be examined to have more than a 90% probability of haplotyping both copies of the gene in an individual. In a particularly preferred embodiment, the nucleotide at each polymorphic or mutant site is identified.

[0070] In a preferred embodiment, a haplotype pair is determined for an individual by identifying the phased sequence of nucleotides at one or more of the polymorphic/mutant sites in each copy of the gene that is present in the individual. In a particularly preferred embodiment, the haplotyping method comprises identifying the phased sequence of nucleotides at each polymorphic/mutant site in each copy of the gene. When haplotyping both copies of the gene, the identifying step is preferably performed with each copy of the gene being placed in separate containers. However, if the two copies are labelled with different tags, or are otherwise separately distinguishable or identifiable, it is possible in some cases to perform the method in the same container. For example, if the first and second copies of the gene are labelled with different first and second fluorescent dyes, respectively, and an allele-specific oligonucleotide labelled with yet a third different fluorescent dye is used to assay the polymorphic/mutant sites, then detecting a combination of the first and third dyes would identify the polymorphism or mutation in the first gene copy, while detecting a combination of the second and third dyes would identify the polymorphism or mutation in the second gene copy.

[0071] In both the genotyping and haplotyping methods, the identity of a nucleotide (or nucleotide pair) at a polymorphic and/or mutant site may be determined by amplifying a target region containing the polymorphic and/or mutant sites directly from one or both copies of the gene, or fragments thereof, and sequencing the amplified regions by conventional methods. It will be readily appreciated by the skilled artisan that only one nucleotide will be detected at a polymorphic or mutant site in individuals who are homozygous at that site, while two different nucleotides will be detected if the individual is heterozygous for that site. The polymorphism or mutation may be identified directly, known as positive-type identification, or by inference, referred to as negative-type identification. For example, where a SNP is known to be guanine and cytosine in a reference population, a site may be positively determined to be either guanine or cytosine for all individuals homozygous at that site, or both guanine and cytosine, if the individual is heterozygous at that site. Alternatively, the site may

-18-

be negatively determined to be not guanine (and thus cytosine/cytosine) or not cytosine (and thus guanine/guanine).

[0072] *Indirect Genotyping Method using Polymorphic and Mutation Sites in Linkage Disequilibrium with a Target Polymorphism or Mutation.* In addition, the identity of the alleles present at any of the novel polymorphic/mutant sites of the invention may be indirectly determined by genotyping other polymorphic/mutant sites in linkage disequilibrium with those sites of interest. As described *supra*, two sites are said to be in linkage disequilibrium if the presence of a particular variant (polymorphism or mutation) at one site is indicative of the presence of another variant at a second site. See, Stevens JC, *Mol. Diag.* 4:309-317 (1999). Polymorphic and mutant sites in linkage disequilibrium with the polymorphic or mutant sites of the invention may be located in regions of the same gene or in other genomic regions. Genotyping of a polymorphic/mutant site in linkage disequilibrium with the novel polymorphic/mutant sites described herein may be performed by, but is not limited to, any of the above-mentioned methods for detecting the identity of the allele at a polymorphic/mutant site.

[0073] *Amplifying a Target Gene Region.* The target regions may be amplified using any oligonucleotide-directed amplification method, including but not limited to polymerase chain reaction (PCR). (U.S. Pat. No. 4,965,188), ligase chain reaction (LCR) (Barany *et al.*, *Proc. Natl. Acad. Sci. USA* 88:189-193 (1991); published PCT patent application WO 90/01069), and oligonucleotide ligation assay (OLA) (Landegren *et al.*, *Science* 241: 1077-1080 (1988)). Oligonucleotides useful as primers or probes in such methods should specifically hybridize to a region of the nucleic acid that contains or is adjacent to the polymorphic/mutant site. Typically, the oligonucleotides are between 10 and 35 nucleotides in length and preferably, between 15 and 30 nucleotides in length. Most preferably, the oligonucleotides are 20 to 25 nucleotides long. The exact length of the oligonucleotide will depend on many factors that are routinely considered and practiced by the skilled artisan.

[0074] Other known nucleic acid amplification procedures may be used to amplify the target region including transcription-based amplification systems (U.S. Pat. No. 5,130,238; EP 329,822; U.S. Pat. No. 5,169,766, published PCT patent application WO 89/06700) and isothermal methods (Walker *et al.*, *Proc. Natl. Acad. Sci. USA* 89: 392-396 (1992)).

[0075] A polymorphism or mutation in the target region may be assayed before or after amplification using one of several hybridization-based methods known in the art. Typically,

allele-specific oligonucleotides are utilized in performing such methods. The allele-specific oligonucleotides may be used as differently labelled probe pairs, with one member of the pair showing a perfect match to one variant of a target sequence and the other member showing a perfect match to a different variant. In some embodiments, more than one polymorphic/mutant site may be detected at once using a set of allele-specific oligonucleotides or oligonucleotide pairs. Preferably, the members of the set have melting temperatures within 5°C, and more preferably within 2°C, of each other when hybridizing to each of the polymorphic or mutant sites being detected.

[0076] *Hybridizing Allele-Specific Oligonucleotide to a Target Gene.* Hybridization of an allele-specific oligonucleotide to a target polynucleotide may be performed with both entities in solution, or such hybridization may be performed when either the oligonucleotide or the target polynucleotide is covalently or noncovalently affixed to a solid support. Attachment may be mediated, for example, by antibody-antigen interactions, poly-L-Lys, streptavidin or avidin-biotin, salt bridges, hydrophobic interactions, chemical linkages, UV cross-linking, baking, *etc.* Allele-specific oligonucleotide may be synthesized directly on the solid support or attached to the solid support subsequent to synthesis. Solid-supports suitable for use in detection methods of the invention include substrates made of silicon, glass, plastic, paper and the like, which may be formed, for example, into wells (as in 96-well plates), slides, sheets, membranes, fibres, chips, dishes, and beads. The solid support may be treated, coated or derivatised to facilitate the immobilization of the allele-specific oligonucleotide or target nucleic acid.

[0077] The genotype or haplotype for the gene of an individual may also be determined by hybridization of a nucleic sample containing one or both copies of the gene to nucleic acid arrays and subarrays such as described in WO 95/11995. The arrays would contain a battery of allele-specific oligonucleotides representing each of the polymorphic or mutant sites to be included in the genotype or haplotype.

[0078] *Determining Population Genotypes and Haplotypes and Correlating them with a Trait.* The present invention provides a method for determining the frequency of a genotype or haplotype in a population. The method comprises determining the genotype or the haplotype for a gene present in each member of the population, wherein the genotype or haplotype comprises the nucleotide pair or nucleotide detected at one or more of the polymorphic sites in the gene and mutations identified in the region, and calculating the

frequency at which the genotype or haplotype is found in the population. The population may be a reference population, a family population, a same sex population, a population group, or a trait population (*e.g.*, a group of individuals exhibiting a trait of interest such as a medical condition or response to a therapeutic treatment).

[0079] In another aspect of the invention, frequency data for genotypes and/or haplotypes found in a reference population are used in a method for identifying an association between a trait and a genotype or a haplotype. The trait may be any detectable phenotype, including but not limited to cancer, susceptibility to a disease or response to a treatment. The method involves obtaining data on the frequency of the genotypes or haplotypes of interest in a reference population and comparing the data to the frequency of the genotypes or haplotypes in a population exhibiting the trait. Frequency data for one or both of the reference and trait populations may be obtained by genotyping or haplotyping each individual in the populations using one of the methods described above. The haplotypes for the trait population may be determined directly or, alternatively, by the predictive genotype to haplotype approach described above.

[0080] In preferred embodiments, the trait is susceptibility to a disease, severity of a disease, the staging of a disease or response to a drug. Such methods have applicability in developing diagnostic tests and therapeutic treatments for all pharmacogenetic applications where there is the potential for an association between a genotype and a treatment outcome, including efficacy measurements, PD measurements, PK measurements and side effect measurements.

[0081] In another embodiment, the frequency data for the reference and/or trait populations are obtained by accessing previously determined frequency data, which may be in written or electronic form. For example, the frequency data may be present in a database that is accessible by a computer. Once the frequency data are obtained, the frequencies of the genotypes or haplotypes of interest in the reference and trait populations are compared. In a preferred embodiment, the frequencies of all genotypes and/or haplotypes observed in the populations are compared. If a particular genotype or haplotype for the gene is more frequent in the trait population than in the reference population at a statistically significant amount, then the trait is predicted to be associated with that genotype or haplotype.

[0082] In a preferred embodiment, the haplotype frequency data for different ethnogeographic groups are examined to determine whether they are consistent with Hardy-

Weinberg equilibrium. Hartl DL *et al.*, *Principles of Population Genomics*, 3rd Ed. (Sinauer Associates, Sunderland, MA, 1997). Hardy-Weinberg equilibrium postulates that the frequency of finding the haplotype pair H_1/H_2 is equal to $P_{H-W}(H_1/H_2) = 2p(H_1)p(H_2)$ if $H_1 \neq H_2$ and $P_{H-W}(H_1/H_2) = p(H_1)p(H_2)$ if $H_1 = H_2$. A statistically significant difference between the observed and expected haplotype frequencies could be due to one or more factors including significant inbreeding in the population group, strong selective pressure on the gene, sampling bias, and/or errors in the genotyping process. If large deviations from Hardy-Weinberg equilibrium are observed in an ethnogeographic group, the number of individuals in that group can be increased to see if the deviation is due to a sampling bias. If a larger sample size does not reduce the difference between observed and expected haplotype pair frequencies, then one may wish to consider haplotyping the individual using a direct haplotyping method such as, for example, CLASPER System™ technology (U.S. Pat. No. 5,866,404), SMD, or allele-specific long-range PCR (Michalotos-Beloin *et al.*, *Nucl. Acids Res.* 24: 4841-4843 (1996)).

[0083] In one embodiment of this method for predicting a haplotype pair, the assigning step involves performing the following analysis. First, each of the possible haplotype pairs is compared to the haplotype pairs in the reference population. Generally, only one of the haplotype pairs in the reference population matches a possible haplotype pair and that pair is assigned to the individual. Occasionally, only one haplotype represented in the reference haplotype pairs is consistent with a possible haplotype pair for an individual, and in such cases the individual is assigned a haplotype pair containing this known haplotype and a new haplotype derived by subtracting the known haplotype from the possible haplotype pair. In rare cases, either no haplotypes in the reference population are consistent with the possible haplotype pairs, or alternatively, multiple reference haplotype pairs are consistent with the possible haplotype pairs. In such cases, the individual is preferably haplotyped using a direct molecular haplotyping method such as, for example, those discussed *supra*.

[0084] In a preferred embodiment, statistical analysis is performed by the use of standard ANOVA tests with a Bonferoni correction and/or a bootstrapping method that simulates the genotype phenotype correlation many times and calculates a significance value. When many polymorphisms and/or mutations are being analyzed, a calculation may be performed to correct for a significant association that might be found by chance. For statistical methods useful in the methods of the present invention, see Bailey NTJ, *Statistical Methods in Biology*,

3rd Edition (Cambridge Univ. Press, Cambridge, 1997); Waterman MS, *Introduction to Computational Biology* (CRC Press, 2000) and *Bioinformatics*, Baxevanis AD & Ouellette BFF, eds. (John Wiley & Sons, Inc., 2001).

[0085] In a preferred embodiment of the method, the trait of interest is a clinical response exhibited by a patient to some therapeutic treatment, for example, response to a drug targeting or to a therapeutic treatment for a medical condition.

[0086] In another embodiment of the invention, a detectable genotype or haplotype that is in linkage disequilibrium with a genotype or haplotype of interest may be used as a surrogate marker. A genotype that is in linkage disequilibrium with another genotype is indicated where a particular genotype or haplotype for a given gene is more frequent in the population that also demonstrates the potential surrogate marker genotype than in the reference population. If the frequency is statistically significant, then the marker genotype is predictive of that genotype or haplotype, and can be used as a surrogate marker.

[0087] *Correlating Subject Genotype or Haplotype to Treatment Response.* In order to deduce a correlation between a clinical response to a treatment and a genotype or haplotype, genotype or haplotype data is obtained on the clinical responses exhibited by a population of individuals who received the treatment, hereinafter the "clinical population". This clinical data may be obtained by analyzing the results of a clinical trial that has already been previously conducted and/or by designing and carrying out one or more new clinical trials.

[0088] It is preferred that the individuals included in the clinical population be graded for the existence of the medical condition of interest. This grading of potential patients could employ a standard physical exam or one or more lab tests. Alternatively, grading of patients could use genotyping or haplotyping for situations where there is a strong correlation between haplotype pair and disease susceptibility or severity.

[0089] The therapeutic treatment of interest is administered to each individual in the trial population, and each individual's response to the treatment is measured using one or more predetermined criteria. It is contemplated that in many cases, the trial population will exhibit a range of responses, and that the investigator may choose more than one responder groups (e.g., low, medium, high) made up by the various responses. In addition, the gene for each individual in the trial population is genotyped and/or haplotyped, which may be done before or after administering the treatment.

[0090] These results are then analyzed to determine if any observed variation in clinical response between polymorphism/mutation groups is statistically significant. Statistical analysis methods, which may be used, are described in Fisher LD & vanBelle G, *Biostatistics: A Methodology for the Health Sciences* (Wiley-Interscience, New York, 1993). This analysis may also include a regression calculation of which polymorphic/mutation sites in the gene contribute most significantly to the differences in phenotype.

[0091] A second method for finding correlations between genotype and haplotype content and clinical responses uses predictive models based on error-minimizing optimization algorithms, one of which is a genetic algorithm. Judson R, *Genetic Algorithms and Their Uses in Chemistry*, in *Reviews in Computational Chemistry, Vol. 10*, Lipkowitz KB & Boyd DB, eds. (VCH Publishers, New York, 1997) pp. 1-73. Simulated annealing (Press *et al.*, *Numerical Recipes in C: The Art of Scientific Computing, Ch. 10* (Cambridge University Press, Cambridge, 1992)), neural networks (Rich E & Knight K, *Artificial Intelligence, 2nd Edition, Ch. 10* (McGraw-Hill, New York, 1991), standard gradient descent methods (Press *et al.*, *Numerical Recipes in C: The Art of Scientific Computing, Ch. 10* (Cambridge University Press, Cambridge, 1992), or other global or local optimization approaches (see discussion in Judson, *supra*) can also be used.

[0092] Correlations may also be analyzed using analysis of variation (ANOVA) techniques to determine how much of the variation in the clinical data is explained by different subsets of the polymorphic and mutant sites in the gene. ANOVA is used to test hypotheses about whether a response variable is caused by or correlates with one or more traits or variables that can be measured (Fisher & vanBelle, *supra*, Ch. 10).

[0093] After the clinical, mutation and polymorphism data have been obtained, correlations between individual response and genotype or haplotype content are created. Correlations may be produced in several ways. In one method, individuals are grouped by their genotype or haplotype (or haplotype pair) (also referred to as a polymorphism/mutation group), and then the averages and standard deviations of clinical responses exhibited by the members of each polymorphism/mutation group are calculated.

[0094] From the analyses described above, the skilled artisan that predicts clinical response as a function of genotype or haplotype content may readily construct a mathematical model. The identification of an association between a clinical response and a genotype or haplotype (or haplotype pair) for the gene may be the basis for designing a diagnostic method to

determine those individuals who will or will not respond to the treatment, or alternatively, will respond at a lower level and thus may require more treatment, *i.e.*, a greater dose of a drug or suffer an adverse reaction. The diagnostic method may take one of several forms: for example, a direct DNA test (*i.e.*, genotyping or haplotyping one or more of the polymorphic/mutant sites in the gene), a serological test, or a physical exam measurement. The only requirement is that there be a good correlation between the diagnostic test results and the underlying genotype or haplotype. In a preferred embodiment, this diagnostic method uses the predictive genotyping/haplotyping method described above.

[0095] *Patient Selection for Therapy Based Upon Polymorphisms and/or Mutations.* The application of genotypes and/or haplotypes that correlate with efficacious drug responses will be used to select patients for therapy of existing diseases. Genotypes and haplotypes that correlate with adverse consequences will be used to either modify how the drug is administered (*e.g.*, dose, schedule or in combination with other drugs) or eliminated as an option.

[0096] *Patient Selection for Prophylactic Therapy Based Upon Polymorphisms and/or Mutations.* The application of genotypes and/or haplotypes that correlate with a predisposition for disease will be used to select patients for preventative therapy.

[0097] *Computer System for Storing or Displaying Polymorphism and Mutation Data.* The invention also provides a computer system for storing and displaying polymorphism and mutation data determined for the gene. The computer system comprises a computer processing unit, a display, and a database containing the polymorphism/mutation data. The polymorphism/mutation data includes the polymorphisms, mutations, the genotypes and the haplotypes identified for a given gene in a reference population. In a preferred embodiment, the computer system is capable of producing a display showing haplotypes organized according to their evolutionary relationships. A computer may implement any or all analytical and mathematical operations involved in practicing the methods of the present invention. In addition, the computer may execute a program that generates views (or screens) displayed on a display device and with which the user can interact to view and analyze large amounts of information relating to the gene and its genomic variation, including chromosome location, gene structure, and gene family, gene expression data, polymorphism data, mutation data, genetic sequence data, and clinical population data (*e.g.*, data on ethnogeographic origin, clinical responses, genotypes, and haplotypes for one or more populations). The

polymorphism and mutation data described herein may be stored as part of a relational database (*e.g.*, an instance of an Oracle database or a set of ASCII flat files). These polymorphism and mutation data may be stored on the computer's hard drive, or may, for example, be stored on a CD-ROM or on one or more other storage devices accessible by the computer. For example, the data may be stored on one or more databases in communication with the computer *via* a network.

[0098] *Nucleic Acid-based Diagnostics.* In another aspect, the invention provides SNP and mutation probes, which are useful in classifying subjects according to their types of genetic variation. The SNP and mutation probes according to the invention are oligonucleotides, which discriminate between SNPs or mutations and the wild-type sequence in conventional allelic discrimination assays. In certain preferred embodiments, the oligonucleotides according to this aspect of the invention are complementary to one allele of the SNP/mutant nucleic acid, but not to any other allele of the SNP/Mutant nucleic acid. Oligonucleotides according to this embodiment of the invention can discriminate between SNPs and mutations in various ways. For example, under stringent hybridization conditions, an oligonucleotide of appropriate length will hybridize to one SNP or mutation, but not to any other. The oligonucleotide may be labelled using a radiolabel or a fluorescent molecular tag. Alternatively, an oligonucleotide of appropriate length can be used as a primer for PCR, wherein the 3' terminal nucleotide is complementary to one allele containing a SNP or mutation, but not to any other allele. In this embodiment, the presence or absence of amplification by PCR determines the haplotype of the SNP or the specific mutation.

[0099] Genomic and cDNA fragments of the invention comprise at least one novel polymorphic site or mutation identified herein, have a length of at least 10 nucleotides, and may range up to the full length of the gene. Preferably, a fragment according to the present invention is between 100 and 3000 nucleotides in length, and more preferably between 200 and 2000 nucleotides in length, and most preferably between 500 and 1000 nucleotides in length.

[0100] *Kits of the Invention.* The invention provides nucleic acid and polypeptide detection kits useful for haplotyping and/or genotyping the genes in an individual. Such kits are useful for classifying individuals for the purpose of classifying individuals. Specifically, the invention encompasses kits for detecting the presence of a polypeptide or nucleic acid corresponding to a marker of the invention in a biological sample, *e.g.*, any tissue or bodily

fluid including, but not limited to, serum, plasma, lymph, cystic fluid, urine, stool, cerebrospinal fluid, ascites fluid or blood, and including biopsy samples of body tissue. For example, the kit can comprise a labelled compound or agent capable of detecting a polypeptide or an mRNA encoding a polypeptide corresponding to a marker of the invention in a biological sample and means for determining the amount of the polypeptide or mRNA in the sample, *e.g.*, an antibody which binds the polypeptide or an oligonucleotide probe which binds to DNA or mRNA encoding the polypeptide. Kits can also include instructions for interpreting the results obtained using the kit.

[0101] In another embodiment, the invention provides a kit comprising at least two genotyping oligonucleotides packaged in separate containers. The kit may also contain other components such as hybridization buffer (where the oligonucleotides are to be used as a probe) packaged in a separate container. Alternatively, where the oligonucleotides are to be used to amplify a target region, the kit may contain, packaged in separate containers, a polymerase and a reaction buffer optimized for primer extension mediated by the polymerase, such as in the case of PCR.

[0102] In a preferred embodiment, such kit may further comprise a DNA sample collecting means. In particular, the genotyping primer composition may comprise at least two sets of allele specific primer pairs. Preferably, the two genotyping oligonucleotides are packaged in separate containers.

[0103] For antibody-based kits, the kit can comprise, *e.g.*, (1) a first antibody, *e.g.*, attached to a solid support, which binds to a polypeptide corresponding to a marker of the invention; and, optionally; (2) a second, different antibody which binds to either the polypeptide or the first antibody and is conjugated to a detectable label.

[0104] For oligonucleotide-based kits, the kit can comprise, *e.g.*, (1) an oligonucleotide, *e.g.*, a detectably-labelled oligonucleotide, which hybridizes to a nucleic acid sequence encoding a polypeptide corresponding to a marker of the invention; or (2) a pair of primers useful for amplifying a nucleic acid molecule corresponding to a marker of the invention.

[0105] The kit can also comprise, *e.g.*, a buffering agent, a preservative or a protein-stabilizing agent. The kit can further comprise components necessary for detecting the detectable-label, *e.g.*, an enzyme or a substrate. The kit can also contain a control sample or a series of control samples, which can be assayed and compared to the test sample. Each component of the kit can be enclosed within an individual container and all of the various

-27-

containers can be within a single package, along with instructions for interpreting the results of the assays performed using the kit.

[0106] *Making Polymorphisms and Mutations of the Invention.* Effects of the polymorphisms and mutations identified herein on gene expression may be investigated by preparing recombinant cells and/or organisms, preferably recombinant animals, containing a polymorphic variant and/or mutation of the gene.

[0107] In one aspect, the present invention includes one or more polynucleotides encoding mutant or polymorphic polypeptides, including degenerate variants thereof. The invention also encompasses allelic variants of the same, that is, naturally occurring alternative forms of the isolated polynucleotides that encode mutant polypeptides that are identical, homologous or related to those encoded by the polynucleotides. Alternatively, non-naturally occurring variants may be produced by mutagenesis techniques or by direct synthesis techniques well known in the art. Accordingly, nucleic acid sequences capable of hybridizing at low stringency with any nucleic acid sequences encoding mutant polypeptide of the present invention are considered to be within the scope of the invention. For example, for a nucleic acid sequence of about 20-40 bases, a typical prehybridization, hybridization, and wash protocol is as follows: (1) prehybridization: incubate nitrocellulose filters containing the denatured target DNA for 3-4 hours at 55°C in 5xDenhardt's solution, 6xSSC (20xSSC consists of 175 g NaCl, 88.2 g sodium citrate in 800 ml H₂O adjusted to pH. 7.0 with 10 N NaOH), 0.1% SDS, and 100 mg/ml denatured salmon sperm DNA, (2) hybridization: incubate filters in prehybridization solution plus probe at 42°C for 14-48 hours, (3) wash; three 15 minutes washes in 6xSSC and 0.1% SDS at room temperature, followed by a final 1-1.5 minutes wash in 6xSSC and 0.1% SDS at 55°C. Other equivalent procedures, *e.g.*, employing organic solvents such as formamide, are well known in the art. Standard stringency conditions are well characterized in standard molecular biology cloning texts. See, for example, Sambrook, Fritsch & Maniatis, *Molecular Cloning A Laboratory Manual, 2nd Ed.*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989); Glover DN, *DNA Cloning, Volumes I and II*, (1985); *Oligonucleotide Synthesis*, Gait MJ, ed. (1984); *Nucleic Acid Hybridization*, Hames BD & Higgins SJ, eds. (1984).

[0108] *Recombinant Expression Vectors.* Another aspect of the invention includes vectors containing one or more nucleic acid sequences encoding a mutant or polymorphic polypeptide. In practicing the present invention, many conventional techniques in molecular

-28-

biology, microbiology and recombinant DNA are used. These techniques are well known and are explained in, e.g., *Current Protocols in Molecular Biology*, Vols. I-III, Ausubel, ed. (1997); Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd Edition. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989); Glover DN, *DNA Cloning: A Practical Approach*, Vols. I and II (1985); *Oligonucleotide Synthesis*, Gait, Ed. (1984); *Nucleic Acid Hybridization*, Hames & Higgins, eds. (1985); *Transcription and Translation*, Hames & Higgins, Eds. (1984); *Animal Cell Culture*, Freshney, ed. (1986); *Immobilized Cells and Enzymes* (IRL Press, 1986); Perbal, *A Practical Guide to Molecular Cloning*; the series *Methods in Enzymol.*, (Academic Press, Inc., 1984); *Gene Transfer Vectors for Mammalian Cells*, Miller & Calos, eds. (Cold Spring Harbor Press, Cold Spring Harbor Laboratory, New York, 1987); and *Methods in Enzymology*, Vols. 154 and 155, Wu & Grossman, and Wu, Eds., respectively.

[0109] For recombinant expression of one or more the polypeptides of the invention, the nucleic acid containing all or a portion of the nucleotide sequence encoding the polypeptide is inserted into an appropriate cloning vector, or an expression vector (*i.e.*, a vector that contains the necessary elements for the transcription and translation of the inserted polypeptide coding sequence) by recombinant DNA techniques well known in the art and as detailed below.

[0110] In general, expression vectors useful in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors that are not technically plasmids, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions. Such viral vectors permit infection of a subject and expression in that subject of a compound. Becker *et al.*, *Meth. Cell Biol.* 43: 161 89 (1994).

[0111] The recombinant expression vectors of the invention comprise a nucleic acid encoding a mutant or polymorphic polypeptide in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression that is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequences in a manner that allows for expression of the

nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell).

[0112] The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, *Gene Expression Technology: Methods In Enzymology* (Academic Press, San Diego, Calif., 1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of polypeptide desired, *etc.* The expression vectors of the invention can be introduced into host cells to thereby produce polypeptides or peptides, including fusion polypeptides, encoded by nucleic acids as described herein (*e.g.*, mutant polypeptides and mutant-derived fusion polypeptides, *etc.*).

[0113] *Mutant and Polymorphic Polypeptide-Expressing Host Cells.* Another aspect of the invention pertains to mutant and polymorphic polypeptide-expressing host cells, which contain a nucleic acid encoding one or more mutant/polymorphic polypeptides of the invention. To prepare a recombinant cell of the invention, the desired isogene may be introduced into a host cell in a vector such that the isogene remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location. In a preferred embodiment, the isogene is introduced into a cell in such a way that it recombines with the endogenous gene present in the cell. Such recombination requires the occurrence of a double recombination event, thereby resulting in the desired gene polymorphism or mutation. Vectors for the introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector or vector construct may be used in the invention. Methods such as electroporation, particle bombardment, calcium phosphate co-precipitation and viral transduction for introducing DNA into cells are known in the art; therefore, the choice of method may lie with the competence and preference of the skilled practitioner.

[0114] The recombinant expression vectors of the invention can be designed for expression of mutant polypeptides in prokaryotic or eukaryotic cells. For example, mutant/polymorphic polypeptides can be expressed in bacterial cells such as *Escherichia coli* (*E. coli*), insect cells

(using baculovirus expression vectors), fungal cells, *e.g.*, yeast, yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods In Enzymology* (Academic Press, San Diego, Calif., 1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase. The SMP2 promoter is useful in the expression of polypeptides in smooth muscle cells, Qian *et al.*, *Endocrinology* 140(4): 1826 (1999).

[0115] Expression of polypeptides in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non fusion polypeptides. Fusion vectors add a number of amino acids to a polypeptide encoded therein, usually to the amino terminus of the recombinant polypeptide. Such fusion vectors typically serve three purposes: (i) to increase expression of recombinant polypeptide; (ii) to increase the solubility of the recombinant polypeptide; and (iii) to aid in the purification of the recombinant polypeptide by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant polypeptide to enable separation of the recombinant polypeptide from the fusion moiety subsequent to purification of the fusion polypeptide. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, *Gene* 67: 31 40 (1988)), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S transferase (GST), maltose E binding polypeptide, or polypeptide A, respectively, to the target recombinant polypeptide.

[0116] Examples of suitable inducible non fusion *E. coli* expression vectors include pTc (Amrann *et al.*, *Gene* 69:301 315 (1988)) and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods In Enzymology* (Academic Press, San Diego, Calif., 1990) pp. 60-89).

[0117] One strategy to maximize recombinant polypeptide expression in *E. coli* is to express the polypeptide in host bacteria with an impaired capacity to proteolytically cleave the recombinant polypeptide. See, *e.g.*, Gottesman, *Gene Expression Technology: Methods In Enzymology* (Academic Press, San Diego, Calif., 1990) 119 128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in the expression host, *e.g.*, *E. coli* (see, *e.g.*, Wada *et al.*, *Nucl. Acids Res.* 20: 2111-2118 (1992)). Such

alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques. In another embodiment, the mutant/polymorphic polypeptide expression vector is a yeast expression vector.

[0118] Examples of vectors for expression in yeast *Saccharomyces cerevisiae* include pYepSec1 (Baldari *et al.*, *EMBO J.* 6: 229 234 (1987)), pMFa (Kurjan & Herskowitz, *Cell* 30: 933 943 (1982)), pJRY88 (Schultz *et al.*, *Gene* 54: 113 123 (1987)), pYES2 (InVitrogen Corporation, San Diego, Calif., USA), and picZ (InVitrogen Corp, San Diego, Calif., USA). Alternatively, mutant polypeptide can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of polypeptides in cultured insect cells (*e.g.*, SF9 cells) include the pAc series (Smith *et al.*, *Mol. Cell. Biol.* 3: 2156 2165 (1983)) and the pVL series (Lucklow & Summers, *Virology* 170: 31 39 (1989)).

[0119] In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, *Nature* 329: 842 846 (1987)) and pMT2PC (Kaufman *et al.*, *EMBO J.* 6: 187 195 (1987)). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, and simian virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see, *e.g.*, Chapters 16 and 17 of Sambrook, *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd Ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989).

[0120] In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue specific regulatory elements are used to express the nucleic acid). Tissue specific regulatory elements are known in the art. Nonlimiting examples of suitable tissue specific promoters include the albumin promoter (liver specific; Pinkert, *et al.*, *Genes Dev.* 1: 268 277 (1987)), lymphoid specific promoters (Calame & Eaton, *Adv. Immunol.* 43: 235 275 (1988)), in particular promoters of T cell receptors (Winoto & Baltimore, *EMBO J.* 8: 729 733 (1989)) and immunoglobulins (Banerji *et al.*, *Cell* 33: 729 740 (1983); Queen & Baltimore, *Cell* 33: 741 748 (1983)), neuron specific promoters (*e.g.*, the neurofilament promoter; Byrne & Ruddle, *Proc. Natl. Acad. Sci. USA* 86: 5473 5477 (1989)), pancreas specific promoters (Edlund *et al.*, *Science* 230: 912 916 (1985)), and mammary gland specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No.

264,166). Developmentally regulated promoters are also encompassed, *e.g.*, the murine *hox* promoters (Kessel & Gruss, *Science* 249: 374 379 (1990)) and the α -fetoprotein promoter (Campes & Tilghman, *Genes Dev.* 3: 537 546 (1989)).

[0121] The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to a mutant polypeptide mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see, *e.g.*, Weintraub *et al.*, "Antisense RNA as a molecular tool for genetic analysis," *Reviews Trends in Genetics*, Vol. 1(1) (1986).

[0122] Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

[0123] A host cell can be any prokaryotic or eukaryotic cell. For example, mutant polypeptide can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

[0124] Vector DNA can be introduced into prokaryotic or eukaryotic cells *via* conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium

chloride co precipitation, DEAE dextran mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989), and other laboratory manuals.

[0125] For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding mutant polypeptide or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

[0126] A host cell that includes a compound of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) recombinant mutant/polymorphic polypeptide. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding mutant/polymorphic polypeptide has been introduced) in a suitable medium such that mutant polypeptide is produced. In another embodiment, the method further comprises the step of isolating mutant/polymorphic polypeptide from the medium or the host cell. Purification of recombinant polypeptides is well known in the art and includes ion exchange purification techniques, or affinity purification techniques, for example with an antibody to the compound. Methods of creating antibodies to the compounds of the present invention are discussed below.

[0127] *Transgenic Animals.* Recombinant organisms, *i.e.*, transgenic animals, expressing a variant gene of the invention are prepared using standard procedures known in the art. Transgenic animals carrying the constructs of the invention can be made by several methods known to those having skill in the art. *See, e.g.*, U.S. Pat. No. 5,610,053 and "The Introduction of Foreign Genes into Mice" and the cited references therein, in: *Recombinant DNA*, Watson JD, Gilman M, Witkowski J & Zoller M, eds. (W.H. Freeman and Company, New York) pp. 254-272. Transgenic animals stably expressing a human isogene and

producing human protein can be used as biological models for studying diseases related to abnormal expression and/or activity, and for screening and assaying various candidate drugs, compounds, and treatment regimens to reduce the symptoms or effects of these diseases.

[0128] *Characterizing Gene Expression Level.* Methods to detect and measure mRNA levels (*i.e.*, gene transcription level) and levels of polypeptide gene expression products (*i.e.*, gene translation level) are well-known in the art and include the use of nucleotide microarrays and polypeptide detection methods involving mass spectrometers, reverse-transcription and amplification and/or antibody detection and quantification techniques. See also, Strachan T & Read A, *Human Molecular Genetics*, 2nd Edition. (John Wiley and Sons, Inc. Publication, New York, 1999)).

[0129] *Determination of Target Gene Transcription.* The determination of the level of the expression product of the gene in a biological sample, *e.g.*, the tissue or body fluids of an individual, may be performed in a variety of ways. The term "biological sample" is intended to include tissues, cells, biological fluids and isolates thereof, isolated from a subject, as well as tissues, cells and fluids present within a subject. Many expression detection methods use isolated RNA. For *in vitro* methods, any RNA isolation technique that does not select against the isolation of mRNA can be utilized for the purification of RNA from cells. See, *e.g.*, Ausubel *et al.*, Ed., *Curr. Prot. Mol. Biol.* (John Wiley & Sons, New York, 1987-1999).

[0130] In one embodiment, the level of the mRNA expression product of the target gene is determined. Methods to measure the level of a specific mRNA are well-known in the art and include Northern blot analysis, reverse transcription PCR and real time quantitative PCR or by hybridization to a oligonucleotide array or microarray. In other more preferred embodiments, the determination of the level of expression may be performed by determination of the level of the protein or polypeptide expression product of the gene in body fluids or tissue samples including but not limited to blood or serum. Large numbers of tissue samples can readily be processed using techniques well-known to those of skill in the art, such as, *e.g.*, the single-step RNA isolation process of U.S. Pat. No. 4,843,155.

[0131] The isolated mRNA can be used in hybridization or amplification assays that include, but are not limited to, Southern or Northern analyses, PCR analyses and probe arrays. One preferred diagnostic method for the detection of mRNA levels involves contacting the isolated mRNA with a nucleic acid molecule (probe) that can hybridize to the mRNA encoded by the gene being detected. The nucleic acid probe can be, *e.g.*, a full-length cDNA, or a

-35-

portion thereof, such as an oligonucleotide of at least 7, 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to an mRNA or genomic DNA encoding a marker of the present invention. Other suitable probes for use in the diagnostic assays of the invention are described herein. Hybridization of an mRNA with the probe indicates that the marker in question is being expressed.

[0132] In one format, the probes are immobilized on a solid surface and the mRNA is contacted with the probes, for example, in an Affymetrix gene chip array (Affymetrix, Calif. USA). A skilled artisan can readily adapt known mRNA detection methods for use in detecting the level of mRNA encoded by the markers of the present invention.

[0133] An alternative method for determining the level of mRNA corresponding to a marker of the present invention in a sample involves the process of nucleic acid amplification, *e.g.*, by RT-PCR (the experimental embodiment set forth in U.S. Pat. No. 4,683,202); ligase chain reaction (Barany *et al.*, *Proc. Natl. Acad. Sci. USA* 88:189-193 (1991)) self-sustained sequence replication (Guatelli *et al.*, *Proc. Natl. Acad. Sci. USA* 87: 1874-1878 (1990)); transcriptional amplification system (Kwoh *et al.*, *Proc. Natl. Acad. Sci. USA* 86: 1173-1177 (1989)); Q-Beta Replicase (Lizardi *et al.*, *Biol. Technology* 6: 1197 (1988)); rolling circle replication (U.S. Pat. No. 5,854,033); or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well-known to those of skill in the art. These detection schemes are especially useful for the detection of the nucleic acid molecules if such molecules are present in very low numbers. As used herein, "amplification primers" are defined as being a pair of nucleic acid molecules that can anneal to 5' or 3' regions of a gene (plus and minus strands, respectively, or vice-versa) and contain a short region in between. In general, amplification primers are from about 10-30 nucleotides in length and flank a region from about 50-200 nucleotides in length.

[0134] Real-time quantitative PCR (RT-PCR) is one way to assess gene expression levels, *e.g.*, of genes of the invention, *e.g.*, those containing SNPs and mutations of interest. The RT-PCR assay utilizes an RNA reverse transcriptase to catalyze the synthesis of a DNA strand from an RNA strand, including an mRNA strand. The resultant DNA may be specifically detected and quantified and this process may be used to determine the levels of specific species of mRNA. One method for doing this is TAQMAN® (PE Applied Biosystems, Foster City, Calif., USA) and exploits the 5' nuclease activity of AMPLITAQ GOLD™ DNA polymerase to cleave a specific form of probe during a PCR reaction. This is referred to as a

TAQMAN™ probe. See Luthra *et al.*, *Am. J. Pathol.* 153: 63-68 (1998); Kuimelis *et al.*, *Nucl. Acids Symp. Ser.* 37: 255-256 (1997); and Mullah *et al.*, *Nucl. Acids Res.* 26(4): 1026-1031 (1998)). During the reaction, cleavage of the probe separates a reporter dye and a quencher dye, resulting in increased fluorescence of the reporter. The accumulation of PCR products is detected directly by monitoring the increase in fluorescence of the reporter dye. Heid *et al.*, *Genome Res.* 6(6): 986-994 (1996)). The higher the starting copy number of nucleic acid target, the sooner a significant increase in fluorescence is observed. See Gibson, Heid & Williams *et al.*, *Genome Res.* 6: 995-1001 (1996).

[0135] Other technologies for measuring the transcriptional state of a cell produce pools of restriction fragments of limited complexity for electrophoretic analysis, such as methods combining double restriction enzyme digestion with phasing primers (see, *e.g.*, EP 0 534858 A1), or methods selecting restriction fragments with sites closest to a defined mRNA end. (See, *e.g.*, Prashar & Weissman, *Proc. Natl. Acad. Sci. USA* 93(2) 659-663 (1996)).

[0136] Other methods statistically sample cDNA pools, such as by sequencing sufficient bases, *e.g.*, 20-50 bases, in each of multiple cDNAs to identify each cDNA, or by sequencing short tags, *e.g.*, 9-10 bases, which are generated at known positions relative to a defined mRNA end pathway pattern. See, *e.g.*, Velculescu, *Science* 270: 484-487 (1995). The cDNA levels in the samples are quantified and the mean, average and standard deviation of each cDNA is determined using by standard statistical means well-known to those of skill in the art. Norman T.J. Bailey, *Statistical Methods In Biology, 3rd Edition* (Cambridge University Press, 1995).

[0137] *Detection of Polypeptides: Immunological Detection Methods.* Expression of the protein encoded by the genes of the invention can be detected by a probe which is detectably labelled, or which can be subsequently labelled. The term "labelled", with regard to the probe or antibody, is intended to encompass direct-labelling of the probe or antibody by coupling, *i.e.*, physically linking, a detectable substance to the probe or antibody, as well as indirect-labelling of the probe or antibody by reactivity with another reagent that is directly-labelled. Examples of indirect labelling include detection of a primary antibody using a fluorescently-labelled secondary antibody and end-labelling of a DNA probe with biotin such that it can be detected with fluorescently-labelled streptavidin. Generally, the probe is an antibody that recognizes the expressed protein. A variety of formats can be employed to determine whether a sample contains a target protein that binds to a given antibody. Immunoassay methods

-37-

useful in the detection of target polypeptides of the present invention include, but are not limited to, *e.g.*, dot blotting, western blotting, protein chips, competitive and non-competitive protein binding assays, immunohistochemistry, enzyme-linked immunosorbant assays (ELISA), fluorescence activated cell sorting (FACS), and others commonly used and widely-described in scientific and patent literature, and many employed commercially. A skilled artisan can readily adapt known protein/antibody detection methods for use in determining whether cells express a marker of the present invention and the relative concentration of that specific polypeptide expression product in blood or other body tissues. Proteins from individuals can be isolated using techniques that are well-known to those of skill in the art. The protein isolation methods employed can, *e.g.*, be such as those described in Harlow & Lane, *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1988)).

[0138] For the production of antibodies to a protein encoded by one of the disclosed genes, various host animals may be immunized by injection with the polypeptide, or a portion thereof. Such host animals may include, but are not limited to, rabbits, mice and rats. Various adjuvants may be used to increase the immunological response, depending on the host species including, but not limited to, Freund's (complete and incomplete), mineral gels, such as aluminium hydroxide; surface active substances, such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet haemocyanin and dinitrophenol; and potentially useful human adjuvants, such as bacille Camette-Guerin (BCG) and *Corynebacterium parvum*.

[0139] Monoclonal antibodies (mAbs), which are homogeneous populations of antibodies to a particular antigen, may be obtained by any technique that provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique of Kohler & Milstein, *Nature* 256: 495-497 (1975); and U.S. Pat. No. 4,376,110; the human B-cell hybridoma technique of Kosbor *et al.*, *Immunol. Today* 4: 72 (1983); Cole *et al.*, *Proc. Natl. Acad. Sci. USA* 80: 2026-2030 (1983); and the EBV-hybridoma technique of Cole *et al.*, *Monoclonal Antibodies and Cancer Therapy* (Alan R. Liss, Inc., 1985) pp. 77-96.

[0140] In addition, techniques developed for the production of "chimaeric antibodies" (see Morrison *et al.*, *Proc. Natl. Acad. Sci. USA* 81: 6851-6855 (1984); Neuberger *et al.*, *Nature* 312: 604-608 (1984); and Takeda *et al.*, *Nature* 314: 452-454 (1985)), by splicing the genes

-38-

from a mouse antibody molecule of appropriate antigen specificity together with genes from a human antibody molecule of appropriate biological activity can be used. A chimaeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable or hypervariable region derived from a murine mAb and a human immunoglobulin constant region.

[0141] Alternatively, techniques described for the production of single chain antibodies (U.S. Pat. No. 4,946,778; Bird, *Science* 242: 423-426 (1988); Huston *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5879-5883 (1988); and Ward *et al.*, *Nature* 334: 544-546 (1989)) can be adapted to produce differentially expressed gene single-chain antibodies.

[0142] Techniques useful for the production of "humanized antibodies" can be adapted to produce antibodies to the proteins, fragments or derivatives thereof. Such techniques are disclosed in U.S. Pat. Nos. 5,932,448; 5,693,762; 5,693,761; 5,585,089; 5,530,101; 5,569,825; 5,625,126; 5,633,425; 5,789,650; 5,661,016; and 5,770,429.

[0143] Antibodies or antibody fragments can be used in methods, such as Western blots or immunofluorescence techniques, to detect the expressed proteins. In such uses, it is generally preferable to immobilize either the antibody or proteins on a solid support. Suitable solid phase supports or carriers include any support capable of binding an antigen or an antibody. Well-known supports or carriers include glass, polystyrene, polypropylene, polyethylene, dextran, nylon, amylases, natural and modified celluloses, polyacrylamides, gabbros and magnetite.

[0144] A useful method, for ease of detection, is the sandwich ELISA, of which a number of variations exist, all of which are intended to be used in the methods and assays of the present invention. As used herein, "sandwich assay" is intended to encompass all variations on the basic two-site technique. Immunofluorescence and EIA techniques are both very well-established in the art. However, other reporter molecules, such as radioisotopes, chemiluminescent or bioluminescent molecules may also be employed. It will be readily apparent to the skilled artisan how to vary the procedure to suit the required use.

[0145] Whole genome monitoring of protein, *i.e.*, the "proteome," can be carried out by constructing a microarray in which binding sites comprise immobilized, preferably monoclonal, antibodies specific to a plurality of protein species encoded by the cell genome. Preferably, antibodies are present for a substantial fraction of the encoded proteins, or at least for those proteins relevant to testing or confirming a biological network model of interest.

As noted above, methods for making monoclonal antibodies are well-known. See, e.g., Harlow & Lane, *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1988)). In a preferred embodiment, monoclonal antibodies are raised against synthetic peptide fragments designed based on genomic sequence of the cell. With such an antibody array, proteins from the cell are contacted to the array and their binding is measured with assays known in the art.

[0146] *Detection of Polypeptides. Two-Dimensional Gel Electrophoresis.* Two-dimensional gel electrophoresis is well-known in the art and typically involves isoelectric focusing along a first dimension followed by SDS-PAGE electrophoresis along a second dimension. See, e.g., Hames *et al.*, *Gel Electrophoresis of Proteins: A Practical Approach* (IRL Press, New York, 1990); Shevchenko *et al.*, *Proc. Natl. Acad. Sci. USA* 93: 14440-14445 (1996); Sagliocco *et al.*, *Yeast* 12: 1519-1533 (1996); and Lander, *Science* 274: 536-539 (1996)).

[0147] *Detection of Polypeptides. Mass Spectroscopy.* The identity as well as expression level of target polypeptide can be determined using mass spectroscopy technique (MS). MS-based analysis methodology is useful for analysis of isolated target polypeptide as well as analysis of target polypeptide in a biological sample. MS formats for use in analyzing a target polypeptide include ionization (I) techniques, such as, but not limited to, matrix assisted laser desorption (MALDI), continuous or pulsed electrospray ionization (ESI) and related methods, such as ionspray or thermospray, and massive cluster impact (MCI). Such ion sources can be matched with detection formats, including linear or non-linear reflectron time of flight (TOF), single or multiple quadrupole, single or multiple magnetic sector Fourier transform ion cyclotron resonance (FTICR), ion trap and combinations thereof such as ion-trap/TOF. For ionization, numerous matrix/wavelength combinations (e.g., matrix assisted laser desorption (MALDI)) or solvent combinations (e.g., ESI) can be employed.

[0148] For mass spectroscopy (MS) analysis, the target polypeptide can be solubilised in an appropriate solution or reagent system. The selection of a solution or reagent system, e.g., an organic or inorganic solvent, will depend on the properties of the target polypeptide and the type of MS performed, and is based on methods well-known in the art. See, e.g., Vorm *et al.*, *Anal. Chem.* 61: 3281 (1994) for MALDI; and Valaskovic *et al.*, *Anal. Chem.* 67: 3802 (1995), for ESI. MS of peptides also is described, e.g., in International PCT Application No. WO 93/24834 and U.S. Pat. No. 5,792,664. A solvent is selected that minimizes the risk that the target polypeptide will be decomposed by the energy introduced for the vaporization

process. A reduced risk of target polypeptide decomposition can be achieved, *e.g.*, by embedding the sample in a matrix. A suitable matrix can be an organic compound such as a sugar, *e.g.*, a pentose or hexose, or a polysaccharide such as cellulose. Such compounds are decomposed thermolytically into CO₂ and H₂O such that no residues are formed that can lead to chemical reactions. The matrix also can be an inorganic compound, such as nitrate of ammonium, which is decomposed essentially without leaving any residue. Use of these and other solvents is known to those of skill in the art. See, *e.g.*, U.S. Pat. No. 5,062,935. Electrospray MS has been described by Fenn *et al.*, *J. Phys. Chem.* 88: 4451-4459 (1984); and PCT Application No. WO 90/14148; and current applications are summarized in review articles. See Smith *et al.*, *Anal. Chem.* 62: 882-89 (1990); and Ardrey, *Spectroscopy* 4: 10-18 (1992).

[0149] The mass of a target polypeptide determined by MS can be compared to the mass of a corresponding known polypeptide. For example, where the target polypeptide is a mutant protein, the corresponding known polypeptide can be the corresponding non-mutant protein, *e.g.*, wild-type protein. With ESI, the determination of molecular weights in femtomole amounts of sample is very accurate due to the presence of multiple ion peaks, all of which can be used for mass calculation. Sub-attomole levels of protein have been detected, *e.g.*, using ESI MS (Valaskovic *et al.*, *Science* 273: 1199-1202 (1996)) and MALDI MS (Li *et al.*, *J. Am. Chem. Soc.* 118: 1662-1663 (1996)).

[0150] *Matrix Assisted Laser Desorption (MALDI)*. The level of the target protein in a biological sample, *e.g.*, body fluid or tissue sample, may be measured by means of mass spectrometric (MS) methods including, but not limited to, those techniques known in the art as matrix-assisted laser desorption/ionization, time-of-flight mass spectrometry (MALDI-TOF-MS) and surfaces enhanced for laser desorption/ionization, time-of-flight mass spectrometry (SELDI-TOF-MS) as further detailed below. Methods for performing MALDI are well-known to those of skill in the art. See, *e.g.*, Juhasz *et al.*, *Analysis, Anal. Chem.* 68: 941-946 (1996), and see also, *e.g.*, U.S. Pat. Nos. 5,777,325; 5,742,049; 5,654,545; 5,641,959; 5,654,545 and 5,760,393 for descriptions of MALDI and delayed extraction protocols. Numerous methods for improving resolution are also known. MALDI-TOF-MS has been described by Hillenkamp *et al.*, *Biological Mass Spectrometry*, Burlingame & McCloskey, eds. (Elsevier Science Publ., Amsterdam, 1990) pp. 49-60.

[0151] A variety of techniques for marker detection using mass spectroscopy can be used. See *Bordeaux Mass Spectrometry Conference Report*, Hillenkamp, Ed., pp. 354-362 (1988); *Bordeaux Mass Spectrometry Conference Report*, Karas & Hillenkamp, Eds., pp. 416-417 (1988); Karas & Hillenkamp, *Anal. Chem.* 60: 2299-2301 (1988); and Karas *et al.*, *Biomed. Environ. Mass Spectrum* 18: 841-843 (1989). The use of laser beams in TOF-MS is shown, *e.g.*, in U.S. Patent Nos. 4,694,167; 4,686,366, 4,295,046 and 5,045,694, which are incorporated herein by reference in their entireties. Other MS techniques allow the successful volatilization of high molecular weight biopolymers, without fragmentation, and have enabled a wide variety of biological macromolecules to be analyzed by mass spectrometry.

[0152] *Surfaces Enhanced for Laser Desorption/Ionization (SELDI)*. Other techniques are used which employ new MS probe element compositions with surfaces that allow the probe element to actively participate in the capture and docking of specific analytes, described as Affinity Mass Spectrometry (AMS). See SELDI patents U.S. Pat. Nos. 5,719,060; 5,894,063; 6,020,208; 6,027,942; 6,124,137; and U.S. Patent application No. U.S. 2003/0003465.

Several types of new MS probe elements have been designed with Surfaces Enhanced for Affinity Capture (SEAC). See Hutchens & Yip, *Rapid Commun. Mass Spectrom.* 7: 576-580 (1993). SEAC probe elements have been used successfully to retrieve and tether different classes of biopolymers, particularly proteins, by exploiting what is known about protein surface structures and biospecific molecular recognition. The immobilized affinity capture devices on the MS probe element surface, *i.e.*, SEAC, determines the location and affinity (specificity) of the analyte for the probe surface, therefore the subsequent analytical MS process is efficient.

[0153] Within the general category of SELDI are three separate subcategories: (1) Surfaces Enhanced for Neat Desorption (SEND), where the probe element surfaces, *i.e.*, sample presenting means, are designed to contain Energy Absorbing Molecules (EAM) instead of "matrix" to facilitate desorption/ionizations of analytes added directly (neat) to the surface; (2) SEAC, where the probe element surfaces, *i.e.*, sample presenting means, are designed to contain chemically defined and/or biologically defined affinity capture devices to facilitate either the specific or non-specific attachment or adsorption (so-called docking or tethering) of analytes to the probe surface, by a variety of mechanisms (mostly non-covalent); and (3) Surfaces Enhanced for Photolabile Attachment and Release (SEPAR), where the probe element surfaces, *i.e.*, sample presenting means, are designed or modified to contain one or

more types of chemically defined cross-linking molecules to serve as covalent docking devices. The chemical specificities determining the type and number of the photolabile molecule attachment points between the SEPAR sample presenting means (*i.e.*, probe element surface) and the analyte (*e.g.*, protein) may involve any one or more of a number of different residues or chemical structures in the analyte (*e.g.*, His, Lys, Arg, Tyr, Phe and Cys residues in the case of proteins and peptides).

[0154] *Functionalizing Polypeptides.* A polypeptide of interest also can be modified to facilitate conjugation to a solid support. A chemical or physical moiety can be incorporate into the polypeptide at an appropriate position. For example, a polypeptide of interest can be modified by adding an appropriate functional group to the carboxyl terminus or amino terminus of the polypeptide, or to an amino acid in the peptide, (*e.g.*, to a reactive side chain, or to the peptide backbone. The artisan will recognize, however, that such a modification, *e.g.*, the incorporation of a biotin moiety, can affect the ability of a particular reagent to interact specifically with the polypeptide and, accordingly, will consider this factor, if relevant, in selecting how best to modify a polypeptide of interest. A naturally-occurring amino acid normally present in the polypeptide also can contain a functional group suitable for conjugating the polypeptide to the solid support. For example, a cysteine residue present in the polypeptide can be used to conjugate the polypeptide to a support containing a sulfhydryl group through a disulfide linkage, *e.g.*, a support having cysteine residues attached thereto. Other bonds that can be formed between two amino acids, include, but are not limited to, *e.g.*, monosulfide bonds between two lanthionine residues, which are non-naturally-occurring amino acids that can be incorporated into a polypeptide; a lactam bond formed by a transamidation reaction between the side chains of an acidic amino acid and a basic amino acid, such as between the γ -carboxyl group of Glu (or alpha carboxyl group of Asp) and the amino group of Lys; or a lactone bond produced, *e.g.*, by a crosslink between the hydroxy group of Ser and the carboxyl group of Glu (or alpha carboxyl group of Asp). Thus, a solid support can be modified to contain a desired amino acid residue, *e.g.*, a Glu residue, and a polypeptide having a Ser residue, particularly a Ser residue at the N-terminus or C-terminus, can be conjugated to the solid support through the formation of a lactone bond. The support need not be modified to contain the particular amino acid, *e.g.*, Glu, where it is desired to form a lactone-like bond with a Ser in the polypeptide, but can be modified,

instead, to contain an accessible carboxyl group, thus providing a function corresponding to the alpha carboxyl group of Glu.

[0155] *Thiol-Reactive Functionalities.* A thiol-reactive functionality is particularly useful for conjugating a polypeptide to a solid support. A thiol-reactive functionality is a chemical group that can rapidly react with a nucleophilic thiol moiety to produce a covalent bond, *e.g.*, a disulfide bond or a thioether bond. A variety of thiol-reactive functionalities are known in the art, including, *e.g.*, haloacetyls, such as iodoacetyl; diazoketones; epoxy ketones, alpha- and beta-unsaturated carbonyls, such as alpha-enones and beta-enones; and other reactive Michael acceptors, such as maleimide; acid halides; benzyl halides; and the like. See Greene & Wuts, *Protective Groups in Organic Synthesis*, 2nd Edition (John Wiley & Sons, 1991).

[0156] If desired, the thiol groups can be blocked with a photocleavable protecting group, which then can be selectively cleaved, *e.g.*, by photolithography, to provide portions of a surface activated for immobilization of a polypeptide of interest. Photocleavable protecting groups are known in the art (see, *e.g.*, published International PCT Application No.

WO 92/10092; and McCray *et al.*, *Ann. Rev. Biophys. Biophys. Chem.* 18: 239-270 (1989)) and can be selectively de-blocked by irradiation of selected areas of the surface using, *e.g.*, a photolithography mask.

[0157] *Linkers.* A polypeptide of interest can be attached directly to a support *via* a linker. Any linkers known to those of skill in the art to be suitable for linking peptides or amino acids to supports, either directly or *via* a spacer, may be used. For example, the polypeptide can be conjugated to a support, such as a bead, through means of a variable spacer. Linkers, include, Rink amide linkers (see, *e.g.*, Rink, *Tetrahedron Lett.* 28: 3787 (1976)); trityl chloride linkers (see, *e.g.*, Leznoff, *Ace Chem. Res.* 11: 327 (1978)); and Merrifield linkers (see, *e.g.*, Bodansky *et al.*, *Peptide Synthesis*, 2nd Edition (Academic Press, New York, 1976)). For example, trityl linkers are known. See, *e.g.*, U.S. Pat. Nos. 5,410,068 and 5,612,474. Amino trityl linkers are also known. See, *e.g.*, U.S. Pat. No. 5,198,531. Other linkers include those that can be incorporated into fusion proteins and expressed in a host cell. Such linkers may be selected amino acids, enzyme substrates or any suitable peptide. The linker may be made, *e.g.*, by appropriate selection of primers when isolating the nucleic acid. Alternatively, they may be added by post-translational modification of the protein of interest. Linkers that are suitable for chemically linking peptides to supports, include disulfide bonds, thioether bonds,

hindered disulfide bonds and covalent bonds between free reactive groups, such as amine and thiol groups.

[0158] *Cleavable Linkers.* A linker can provide a reversible linkage such that it is cleaved under the select conditions. In particular, selectively cleavable linkers, including photocleavable linkers (see U.S. Pat. No. 5,643,722), acid cleavable linkers (see Fattom *et al.*, *Infect. Immun.* 60: 584-589 (1992)), acid-labile linkers (see Welhöner *et al.*, *J. Biol. Chem.* 266: 4309-4314 (1991)) and heat sensitive linkers are useful. A linkage can be, *e.g.*, a disulfide bond, which is chemically cleavable by mercaptoethanol or dithioerythrol; a biotin/streptavidin linkage, which can be photocleavable; a heterobifunctional derivative of a trityl ether group, which can be cleaved by exposure to acidic conditions or under conditions of MS (see Köster *et al.*, *Tetrahedron Lett.* 31: 7095 (1990)); a levulinyl-mediated linkage, which can be cleaved under almost neutral conditions with a hydrazinium/acetate buffer; an arginine-arginine or a lysine-lysine bond, either of which can be cleaved by an endopeptidase, such as trypsin; a pyrophosphate bond, which can be cleaved by a pyrophosphatase; or a ribonucleotide bond, which can be cleaved using a ribonuclease or by exposure to alkali condition. A photolabile cross-linker, such as 3-amino-(2-nitrophenyl)propionic acid can be employed as a means for cleaving a polypeptide from a solid support. Brown *et al.*, *Mol Divers*, pp. 4-12 (1995); Rothschild *et al.*, *Nucl. Acids. Res.* 24: 351-66 (1996); and U.S. Pat. No. 5,643,722. Other linkers include RNA linkers that are cleavable by ribozymes and other RNA enzymes and linkers, such as the various domains, such as CH₁, CH₂ and CH₃, from the constant region of human IgG1. See, Batra *et al.*, *Mol Immunol* 30: 379-396 (1993).

[0159] Combinations of any linkers are also contemplated herein. For example, a linker that is cleavable under MS conditions, such as a silyl linkage or photocleavable linkage, can be combined with a linker, such as an avidin biotin linkage, that is not cleaved under these conditions, but may be cleaved under other conditions. Acid-labile linkers are particularly useful chemically cleavable linkers for mass spectrometry, especially for MALDI-TOF, because the acid labile bond is cleaved during conditioning of the target polypeptide upon addition of a 3-HPA matrix solution. The acid labile bond can be introduced as a separate linker group, *e.g.*, an acid labile trityl group, or can be incorporated in a synthetic linker by introducing one or more silyl bridges using diisopropylsilyl, thereby forming a diisopropylsilyl linkage between the polypeptide and the solid support. The diisopropylsilyl linkage can be cleaved using mildly acidic conditions, such as 1.5%

trifluoroacetic acid (TFA) or 3-HPA/1% TFA MALDI-TOF matrix solution. Methods for the preparation of diisopropylsilyl linkages and analogues thereof are well-known in the art.

See, *e.g.*, Saha *et al.*, *J. Org. Chem.* 58: 7827-7831 (1993).

[0160] *Use of a Pin Tool to Immobilize a Polypeptide.* The immobilization of a polypeptide of interest to a solid support using a pin tool can be particularly advantageous. Pin tools include those disclosed herein or otherwise known in the art. See, *e.g.*, U.S. Application Serial Nos. 08/786,988 and 08/787,639; and International PCT Application No. WO 98/20166.

[0161] A pin tool in an array, *e.g.*, a 4×4 array, can be applied to wells containing polypeptides of interest. Where the pin tool has a functional group attached to each pin tip, or a solid support, *e.g.*, functionalized beads or paramagnetic beads are attached to each pin, the polypeptides in a well can be captured (1 pmol capacity). During the capture step, the pins can be kept in motion (vertical, 1-2 mm travel) to increase the efficiency of the capture.

Where a reaction, such as an *in vitro* transcription is being performed in the wells, movement of the pins can increase efficiency of the reaction. Further immobilization can result by applying an electrical field to the pin tool. When a voltage is applied to the pin tool, the polypeptides are attracted to the anode or the cathode, depending on their net charge.

[0162] For more specificity, the pin tool (with or without voltage) can be modified to have conjugated thereto a reagent specific for the polypeptide of interest, such that only the polypeptides of interest are bound by the pins. For example, the pins can have nickel ions attached, such that only polypeptides containing a polyhistidine sequence are bound. Similarly, the pins can have antibodies specific for a target polypeptide attached thereto, or to beads that, in turn, are attached to the pins, such that only the target polypeptides, which contain the epitope recognized by the antibody, are bound by the pins.

[0163] Captured polypeptides can be analyzed by a variety of means including, *e.g.*, spectrometric techniques, such as UV/VIS, IR, fluorescence, chemiluminescence, NMR spectroscopy, MS or other methods known in the art, or combinations thereof. If conditions preclude direct analysis of captured polypeptides, the polypeptides can be released or transferred from the pins, under conditions such that the advantages of sample concentration are not lost. Accordingly, the polypeptides can be removed from the pins using a minimal volume of eluent, and without any loss of sample. Where the polypeptides are bound to the

-46-

beads attached to the pins, the beads containing the polypeptides can be removed from the pins and measurements made directly from the beads.

[0164] Pin tools can be useful for immobilizing polypeptides of interest in spatially addressable manner on an array. Such spatially addressable or pre-addressable arrays are useful in a variety of processes, including, for example, quality control and amino acid sequencing diagnostics. The pin tools described in the U.S. Application Nos. 08/786,988 and 08/787,639 and International PCT Application No. WO 98/20166 are serial and parallel dispensing tools that can be employed to generate multi-element arrays of polypeptides on a surface of the solid support. The array surface can be flat, with beads or geometrically altered to include wells, which can contain beads. In addition, MS geometries can be adapted for accommodating a pin tool apparatus.

[0165] *Other Aspects of the Biological State.* In various embodiments of the invention, aspects of the biological activity state, or mixed aspects can be measured in order to obtain drug and pathway responses. The activities of proteins relevant to the characterization of cell function can be measured, and embodiments of this invention can be based on such measurements. Activity measurements can be performed by any functional, biochemical or physical means appropriate to the particular activity being characterized. Where the activity involves a chemical transformation, the cellular protein can be contacted with natural substrates, and the rate of transformation measured. Where the activity involves association in multimeric units, *e.g.*, association of an activated DNA binding complex with DNA, the amount of associated protein or secondary consequences of the association, such as amounts of mRNA transcribed, can be measured. Also, where only a functional activity is known, *e.g.*, as in cell cycle control, performance of the function can be observed. However known and measured, the changes in protein activities form the response data analyzed by the methods of this invention. In alternative and non-limiting embodiments, response data may be formed of mixed aspects of the biological state of a cell. Response data can be constructed from, *e.g.*, changes in certain mRNA abundances, changes in certain protein abundances and changes in certain protein activities.

[0166] The following EXAMPLES are presented in order to more fully illustrate the preferred embodiments of the invention. These EXAMPLES should in no way be construed as limiting the scope of the invention, as defined by the appended claims.

EXAMPLE I

BIOINFORMATICS ANALYSIS OF FLT-4 MUTATIONS

[0168] *Identification of FLT-4 Mutations in Colon Cancer.* To determine FLT-4 mutations in association with cancer, DHPLC analysis (Lilleberg SL, *Curr. Opin. Drug Discov. Devel.*, 6(2): 237-52 (March 2003)) was conducted on test samples derived from human tissues, *e.g.*, colon cancer (*a.k.a.*, colorectal cancer), as summarized in TABLE 3 below. Specifically, blood tissue samples were analyzed from colon cancer patients. Six (6) FLT-4 missense mutations were identified as detailed below in TABLE 3 (see also, TABLE 1 and TABLE 2, *supra*).

TABLE 3
FLT-4 Mutations in Cancer Patients

<u>Mutation</u>	<u>Amino Acid Change</u>	<u>Amino Acid Position</u>	<u>Identifier</u>
AAC>GAC	N>D	149	N149D
ATG>GTG	M>V	963	M963V
GAT>GAG	D>E	970	D970E
Ins 60bp	Insert	1123-1143	E1123-A1143
CCG>TCG	P>S	1114	P1114S

[0169] *Computational Analysis of FLT-4 Mutations.* The FLT-4 mutations identified in human cancer were analyzed using computational analysis tools to determine the effect(s) of these mutations on FLT-4 function.

-48-

[0170] *Comparison of Known FLT-4 Mutations and SNPs with the FLT-4 Mutations and SNPs of the Present Invention.* As summarized in TABLE 4, there were fifteen (15) coding SNPs within the SNP database for FLT-4 (dbSNP:

<http://www.ncbi.nlm.nih.gov/SNP/index.html>). None of the mutations in TABLE 1 matches known non-synonymous SNPs. N149D has been identified as a SNP previously.

TABLE 4
Select FLT-4 SNPs

<u>cSNP</u>	<u>Allele 1</u> <u>frequency</u>	<u>Allele 2</u> <u>frequency</u>	<u>RefSNP</u>
R1189C	C	T	rs744282
E1164D	G	C	rs1049080
P1066P	C	T	rs11949194
H890Q	C: 0.65	G: 0.35	rs448012
T494A	A: 0.85	G: 0.15	rs307826
Y448Y	C	T	rs3736062
Q285Q	G	A	rs1049077
L250L	G	C	rs1049076
L169L	G: 0.917	T: 0.083	rs3736061
N149D	445:A	G	Evans AL <i>et al.</i> , <i>J. Med. Genet.</i> 40(9):697-703, 2003
Q890H	2670:G	C	Evans AL <i>et al.</i> , <i>J. Med. Genet.</i> 40(9):697-703, 2003
P954S	2860:C	T	Evans AL <i>et al.</i> , <i>J. Med. Genet.</i> 40(9):697-703, 2003
P1008L	3023:C	T	Evans AL <i>et al.</i> , <i>J. Med. Genet.</i> 40(9):697-703, 2003
R1146H	3437:G	A	Evans AL <i>et al.</i> , <i>J. Med. Genet.</i> 40(9):697-703, 2003
R1324L	3971:G	T	Evans AL <i>et al.</i> , <i>J. Med. Genet.</i> 40(9):697-703, 2003

-49-

[0171] There are 12 types of FLT-4 mutations listed within the Human Gene Mutation Database (HGMD; Krawczak *et al.*, *Trends Genet.*, 13(3):121-2 (1997); Cooper *et al.*, *Hum. Genet.*, 98(5):629 (1996)) as summarized below in TABLE 5. None of the mutations within this database match any of the FLT-4 mutations of the present invention.

TABLE 5
FLT-4 Mutations in Disease

<u>Codon</u>	<u>Nucleotide</u>	<u>Amino acid</u>	<u>Phenotype</u>
G854S	cGGC-AGC	Gly-Ser	Lymphoedema, primary
G857R	cGGG-AGG	Gly-Arg	Lymphoedema, primary
A915P	gGCG-CCG	Ala-Pro	Lymphoedema, primary
C916W	TGCa-TGG	Cys-Trp	Lymphoedema, primary
G933R	cGGC-CGC	Gly-Arg	Lymphoedema, primary
H1035R	CAC-CGC	His-Arg	Lymphoedema, primary
R1041Q	CGG-CAG	Arg-Gln	Lymphoedema, primary
R1041P	CGG-CCG	Arg-Pro	Lymphoedema, primary
R1041W	tCGG-TGG	Arg-Trp	Lymphoedema, primary
L1044P	CTG-CCG	Leu-Pro	Lymphoedema, primary
P1114L	CCG-CTG	Pro-Leu	Lymphoedema, primary
P1137L	CCG-CTG	Pro-Leu	Lymphoedema, primary

[0172] *Analysis of the Effect of FLT-4 Mutations on FLT-4 Protein Domain Structure and Function. Pfam Analysis of the Potential Effect of the FLT-4 mutations on FLT-4 Protein Domain Structure* The effect of the FLT-4 mutations on the protein domain structure of FLT-4 was analyzed using the Pfam computational analysis tool. Pfam is a large collection of multiple sequence alignments and hidden Markov models covering many common protein families based on the Swissprot 44.5 and SP-TrEMBL 27.5 protein sequence databases. A search using Pfam (Bateman *et al.*, *Nucl. Acids Res.*, Database Issue 32:D138-D141 (2004)) showed that FLT-4 contained ten (10) domains as summarized in TABLE 6.

TABLE 6
Pfam Analysis of FLT-4 Domain Structure

<u>Model</u>	<u>Seq-from</u>	<u>Seq-to</u>	<u>Description</u>
I-set	230	328	Immunoglobulin I-set domain
V-set	231	328	Immunoglobulin V-set domain
ig	245	312	Immunoglobulin domain
ig	349	398	Immunoglobulin domain
ig	571	655	Immunoglobulin domain
V-set	677	766	Immunoglobulin V-set domain
I-set	678	766	Immunoglobulin I-set domain
ig	692	750	Immunoglobulin domain
Pkinase	845	1169	Protein kinase domain
Pkinase_Tyr	845	1169	Protein tyrosine kinase

-50-

[0173] Sequence alignments of the wild-type human FLT-4 polypeptide sequence with the Pfam domain polypeptide sequences are summarized below in TABLE 7 through TABLE 16. M963, D970 and P1114 were observed as located in the protein tyrosine kinase domain. Sequence alignment of the wild-type human FLT-4 polypeptide sequence with the Pfam immunoglobulin I-set domain sequence is summarized below in TABLE 7. An immunoglobulin I-set domain was identified in the wild-type FLT-4 polypeptide at amino acid position 230 to 328 (TABLE 7; score 15.1, E = 4.6e-05).

TABLE 7
Sequence Alignment Comparison of Human FLT-4 from AA 230 to AA 328 with Pfam
Model of Immunoglobulin I-set Domain

```

          *->PkFtqkp.kdveVqeGesarFeCkVtGt..PePtVsWfKdGqplssv
              +  |++++|  +||  +++  |++ + +  |  |  |++
FLT-4   230   YDIQLLPzKSLELLVGEKLVLNCTVWAEfnSGVTFDWDYPGKQAERG 276

          1.lpsrrfkvtYdgGsgtytLtIsnvqpdDsGkYtCvAtNsaGeaeasae
              |+++ + |+++  + ||| ||  |  |  |++|  |  +  | |
FLT-4   277   KwVPERRSQQTHTE--LSSILTIHNVSQHDLSYVCKANNGIQRFRESTE 324

          LtVq<-*
          ++|
FLT-4   325   VIVH      328

```

SEQ ID NO:9

-51-

[0174] Sequence alignment of the wild-type human FLT-4 polypeptide sequence with the Pfam immunoglobulin V-set domain sequence is summarized below in TABLE 8. An immunoglobulin V-set domain was identified in the wild-type FLT-4 polypeptide at amino acid position 231 to 328 (TABLE 8; score 22.7, E = 0.00017).

TABLE 8
Sequence Alignment Comparison of Human FLT-4 from AA 231 to AA 328 with Pfam Model of Immunoglobulin V-set Domain

```

          *->qsvvtqepprLLLASLfwptsvtvaeggsvtLpCtysfddsssssgt
          ++++      +      |+      +|+++ |+|++      + +|+
FLT-4   231   DIQLLP--RK-----SLELLVGEKLVLNCTVW---AEFNSGV 262

          gsvyWyrqqppGekpeliisfytyyvspsegngtvserfkgrvrfsgnps
          ++ | +      + |++++|+| + ++      +
FLT-4   263 -TFDWDY-PGK-----QAERGKWVPERRSQQTHTELSSI 294

          kgdfsLtIsnlqlsDsGtYyCa.....gsgtrLtVl<-*
          |||+|+++ | | | | +++ + ++|+++|+
FLT-4   295 ----LTIHNVSQHDLSYVCKanngiqrfRESTEVIVH      328

```

SEQ ID NO:10

[0175] Sequence alignment of the wild-type human FLT-4 polypeptide sequence with the Pfam immunoglobulin domain sequence is summarized below in TABLE 9. An immunoglobulin domain was identified in the wild-type FLT-4 polypeptide at amino acid position 245 to 312 (TABLE 9; score 29.7, E = 9e-06).

TABLE 9
Sequence Alignment Comparison of Human FLT-4 from AA 245 to AA 312 with Pfam Model of Immunoglobulin Domain

```

          *->GesvtLtCsv...gppdvtvtWlrngk.le.gvlvtssenngdstyq
          ||++ |+|+| +++++ ||+ | ||+ |+| ++++++++
FLT-4   245   GEKLVLNCTVwaeFNSGVTFDWDYPGKqAErG---KWVPERRSQQTH 288

          t.....LtitnvtpeDsgtYtCvv<-*
          |+ ++ |||+|+++ | |+| | +
FLT-4   289 TelssiLTIHNVSQHDLSYVCKA      312

```

SEQ ID NO:11

-52-

[0176] Sequence alignment of the wild-type human FLT-4 polypeptide sequence with the Pfam immunoglobulin domain sequence is summarized below in TABLE 10. An immunoglobulin domain was identified in the wild-type FLT-4 polypeptide at amino acid position 349 to 398 (TABLE 10; score 13.8, E = 0.0038).

TABLE 10
Sequence Alignment Comparison of Human FLT-4 from AA 349 to AA 398 with Pfam Model of Immunoglobulin Domain

```

          *->GesvtLtCsv.gppdvtvtWlrngk.le.gvlvtssenngdstyqtL
          | |+|+   + |++++|+++|| | + ++ +   |
FLT-4    349    DELVKLPVKLaAYPPPEFQWYKDGKaLSgRHSPHA-----L 384

          titnvtpeDsgtYt<-*
          ++++||+  +||||
FLT-4    385 VLKEVTEASTGTYT      398

```

SEQ ID NO:12

[0177] Sequence alignment of the wild-type human FLT-4 polypeptide sequence with the Pfam immunoglobulin domain sequence is summarized below in TABLE 11. An immunoglobulin domain was identified in the wild-type FLT-4 polypeptide at amino acid position 571 to 655 (TABLE 11; score 15.6, E = 0.15).

TABLE 11
Sequence Alignment Comparison of Human FLT-4 from AA 571 to AA 655 with Pfam Model of Immunoglobulin Domain

```

          *->GesvtLtCsv.gppdvtvtWlrngk.le.gvlvtssenngdstyqt.
          |++| |+|++++ + +++|+| + ++ ++ ++   +
FLT-4    571    GQPVLLSCQAdSYKYEHLRWYRLNLsTLhDAHGNPLL-----Ld 609

          .....LtitnvtpeDsgtYtCvv<-*
          ++ +   ++   + ++ ++ ++ + +|+| +| ||++| | |+|
FLT-4    610 cknvhlfatplaasleevapgarhatlsLSIPRVAPEHEGHYVCEV      655

```

SEQ ID NO:13

-53-

[0178] Sequence alignment of the wild-type human FLT-4 polypeptide sequence with the Pfam immunoglobulin V-set domain sequence is summarized below in TABLE 12. A protein kinase domain was identified in the wild-type FLT-4 polypeptide at amino acid position 677 to 766 (TABLE 12; score 18.8, E = 0.00045).

TABLE 12
Sequence Alignment Comparison of Human FLT-4 from AA 677 to AA 766 with Pfam
Model of Immunoglobulin V-Set Domain

```

      *->qsvvtqepprLLLASLfwptsvtvaegsvtLpCtysfddsssssgt
      +  +||  +          + |++ +|++++| +  +  +
FLT-4  677  APRLTQ-NLT-----DLLVNVSDSLEMQLVA-GAH----AP 707

      gsvyWyrq.qppGekpeliisfytyyvspsegngtvserfkgrvrfsgnp
      |+ ||++++ +          + |  | ++ ++
FLT-4  708  -SIVWYKDeRLL-----EEKSGVDLADSN 730

      skgdfsLtIsnqlsDsGtYyCa.....gsgtrLtVl<-*
      |+|++++ +| | |  + ++  |+  + |+
FLT-4  731  Q----KLSIQRVREEDAGRYLCSvcnakgcvNSSASVAVE 766

```

SEQ ID NO:14

[0179] Sequence alignment of the wild-type human FLT-4 polypeptide sequence with the Pfam immunoglobulin I-set domain sequence is summarized below in TABLE 13. An immunoglobulin I-set domain was identified in the wild-type FLT-4 polypeptide at amino acid position 678 to 766 (TABLE 13; score 52.1, E = 1.6e-12).

TABLE 13
Sequence Alignment Comparison of Human FLT-4 from AA 678 to AA 766 with Pfam
Model of Immunoglobulin I-Set Domain

```

      *  ->PkFtqpkdveVqeGesarFeCkVtGtPePtVsWfKdGqplssvllp
      |+ || + |+ |++ +|++++| | |  |++ |+|| + | +
FLT-4  678  PRLTQNLTDLLVNVSDSLEMQLVAGAHAPSIVWYKDERLLEE---- 720

      srrfkvtYdgGsgtytLtIsnvpqdDsGkYtCvAtNsaGeaeasaeLtVq
      + |+  +  +++| |++|  +|+| | | + |  + +||++ |+
FLT-4  721  --KSGVDLAD--SNQKLSIQRVREEDAGRYLCSVCNAKGCVNSSASVAVE 766

```

SEQ ID NO:15

[0180] Sequence alignment of the wild-type human FLT-4 polypeptide sequence with the immunoglobulin domain sequence is summarized below in TABLE 14. An immunoglobulin domain was identified in the wild-type FLT-4 polypeptide at amino acid position 692 to 750 (TABLE 14; score 33.7, E = 5.6e-07).

-54-

TABLE 14
Sequence Alignment Comparison of Human FLT-4 from AA 692 to AA 750 with Pfam
Model of Immunoglobulin Domain

```

          *->GesvtLtCsv.gppdvtvtWlrngk.le.gvltssennngdstyqtL
              +|+ ++| ! |   +++ |+++  ||+  |+  +| +   |
FLT-4   692   SDSLEMQCLVaGAHAPSIVWYKDERILEeKSGVDLADSNQK-----L 733

          titnvtpeDsgtYtCvv<-*
          +|++|++||+| | | |
FLT-4   734 SIQRVREEDAGRYLCSV      750

```

SEQ ID NO:16

```
*->yelleklGsGsfGtV.....yakakhkgtgkiVAvKiikkrkekskdd
+++|++++|| |++|+ + + + ++ +++++||| + |+++ +++
FLT-4      845    LHLGRVLGYGAFGKVveasaFGIHKSSCDTVAVKML-KEGATASEH 890

ktarrrrEiqilkrL.sHpNivrlidvfedkdhly.lVlEyceGLrllld
+++ |++|| +++++| |+| |+++++ ++ + ++ |+| |
FLT-4      891    RALMS--ELKILIHlgNHLNVVNLLGACTKPQGPLmVIVEFCKYG----N 934

LfdyIs..d.....
|+ |+| + ++ ++ ++++++++ + + + ++++++++ +
FLT-4      935    LSNFLRakRdafspcaekspeqrgrframvelarldrrrpqssdrvlfar 984

.....trrkgp11.fsedeakkialQilrgleYlHsnIgi
++++++ ++ +++++ ++ +|+++++++ ++ |++|+|+| |+ ++
FLT-4      985    fskteggarraspdqEAEDLWLSpLTMEDLVCYSFQVARGMEFLASR-KC 1033

iHRDLKPeNiLldengvvdaflKiaDFGlAkkls.t.sss.lttfvGTpw
|||| | |||+| +|| |++||||| + +++ + +++ + +
FLT-4      1034   IHRDLAARNILLSESDVV---KICDFGLARDIYkDpDYVrKGSARLPk 1079

YmAPEvllgngygtkvDvWSlGvilylelt.GkpPFsgdevqdsldqlql
+|||| + ++ |+| +|||+|++|+| ++ | |++| ++
FLT-4      1080   WMAPESIF-DKVYTTQSDVWSFGVLLWEIFSIGASPYPGV-----QINEE 1123

irkilgpplfdspewdsisysefrrlpnlseeakdLikkcLnkDPskRl
+++ |+ ++ || + ++ + ++ +|+ ||+ |+
FLT-4      1124   FCQLRDGTRMRAPELAT-----PAIRRIMLCNWSGDPKARP 1160

taeeilqHPwf<-*
+ |+++ +
FLT-4      1161   AFSELVEI--L 1169
```

SEQ ID NO:17

```
*->lklgkklGeGaFGeVykgTlkg.sgegtkikVAVKtLkeigasseei
|+|++|| | |||+|+++ ++| +++ + +|||+|| | |+ +|
FLT-4      845    LHLGRVLGYGAFGKVVEASAFGiHKGSSCDTVAVKMLKE-GATASE- 889

gredFlrEAsimkkIgdHpNiVrLlGvctkegePgpgglyiVtEymegGd
++++++|++|+ ++|+| |++| |||+||| |+++| |++++|++++|+
FLT-4      890 -HRALMSELKILIHIGNHLNVVNLLGACTKPQGP----LMVIVEFCKYGN 934

LldfLrkhr.....
|+| ||++|+ ++ ++++++++ + + + ++++++++ +
FLT-4      935 LSNFLRAKRdafspcaekspeqrgrframvelarldrrrpqssdrvlfar 984

.....egrpLtIkdLlsfalQiAkGMeyLesknfv
++++++ ++ +++++ ++ + +|||++|+++++|+|+|||+|+|++++
FLT-4      985 fskteggarraspdqeaediWLSPLTMEDLVCYSFQVARGMEFLASRKCI 1034

HRDLAARNcLVsenlvVKIsDFGLaRdiynddyYvrkkggkPlvkWmAP
|||||++|++|+ |||+||| |||++| +| + |+++| | |||
FLT-4      1035 HRDLAARNILLSESDVVKICDFGLARDIYKDPPDY-VRKGSARLPLKWMAP 1083

EsIkygkFtskSDVWSFGVllWEiftlGeqPFYPgm.sneevlellyedG
|++++++|++| ||| ||| |||++| |||++| |||++++| +|
FLT-4      1084 ESIFDKVYTTQSDVWSFGVllWEIfSLGASP-YPGVqINEEFQCRL-RDG 1131

yRLprPenCPdeIYdlMlqCwaedPedRPtFselperL<-*
|+ +|}++++ + ++|| | | |++ | | ||| |||+|
FLT-4      1132 TRMRAPELATPAIRRIMLNCWSGDPKARPASFSELVEIL 1169
```

SEO ID NO:18

-57-

[0183] *Analysis of the Potential Effect of FLT-4 mutations on FLT-4 Protein Regulatory Sites. NetPhos Analysis of the Effect of FLT-4 mutations on FLT-4 Protein Phosphorylation.* Potential FLT-4 phosphorylation sites were identified by computational analysis using the NetPhos computational analysis tool. NetPhos produces neural network predictions for serine, threonine and tyrosine phosphorylation sites in eukaryotic proteins (Blom *et al.*, *J. Mol. Biol.*, 294(5): 1351-1362, 1999). Potential FLT-4 phosphorylation sites predicted by NetPhos are summarized below in TABLE 17. To be considered a potential phosphorylation site a threshold score of 0.5 was required. The predicted phosphorylation sites marked in bold text represent possible mutation interference.

TABLE 17
FLT-4 Phosphorylation Sites Predicted by NetPhos

<u>Phosphorylation</u>	<u>Amino Acid Position</u>
Serine	46, 48, 50, 75, 128, 173, 174, 239, 284, 301, 322, 334, 376, 380, 431, 439, 565, 582, 623, 660, 694, 722, 734, 773, 814, 871, 947, 953, 976, 977, 986, 996, 1030, 1046, 1073, 1085, 1113 , 1191, 1193, 1194, 1200, 1217, 1220, 1225, 1249, 1250, 1268, 1281, 1289, 1295
Threonine	43, 70, 78, 287, 323, 464, 493, 494, 506, 681, 1093, 1132, 1254, 1263, 1264
Tyrosine	109, 129, 230, 268, 308, 435, 583, 585, 651, 669, 812, 830, 932, 1063, 1068

[0184] *PROSITE Analysis of the Potential Effect of FLT-4 mutations on Other FLT-4 Protein Regulatory Sites.* The effect of the FLT-4 mutations on other protein regulatory sites was analyzed using the PROSITE computational analysis tool. PROSITE is a database of protein families and domains. It consists of biologically significant sites, patterns and profiles that help to reliably identify to which known protein family (if any) a new sequence belongs as well as to identify potential sites for protein modification (Hulo N. *et al.*, *Nucl. Acids Res.*, 32:D134-D137 (2004); Sigrist C.J.A. *et al.*, *Brief Bioinform.*, 3:265-274 (2002); Gattiker A. *et al.*, *Applied Bioinformatics*, 1:107-108 (2002)). Other potential sites for protein modification of FLT-4 polypeptide as predicted by ProSite analysis are summarized below in TABLE 18. Amino acid positions highlighted in bold text represent possible interference by mutations. P1114 was observed as close to a potential myristoylation site.

TABLE 18
Potential FLT-4 protein modification sites predicted by PROSITE

<u>Function</u>	<u>Amino Acid Position</u>
PS00107 PROTEIN_KINASE_ATP (pattern) Protein kinases ATP-binding region signature	851-879
PS00240 RECEPTOR_TYR_KIN_III (pattern) Receptor tyrosine kinase class III signature	904-917
PS00109 PROTEIN_KINASE_TYR (pattern) Tyrosine protein kinases specific active-site	1033-1045
PS00008 MYRISTYL (pattern) N-myristoylation site	4-9; 16-21; 20-25; 107-112; 122-127; 164-169; 190-195; 306-311; 407-412; 630-635; 723-728; 755-760; 866-871; 884-889; 904-909; 933-938; 991-996; 1117-1122; 1267-1272; 1276-1281
PS00001 ASN_GLYCOSYLATION (pattern) N-glycosylation site	33-36; 104-107; 166-169; 251-254; 299-302; 411-414; 515-518; 527-530; 594-597; 683-686; 690-693; 758-761; 934-937; 1271-1274
PS00005 PKC_PHOSPHO_SITE (pattern) Protein kinase C phosphorylation site	50-52; 168-170; 376-378; 473-475; 537-539; 582-584; 977-979; 1030-1032; 1073-1075; 1249-1251; 1264-1266
PS00006 CK2_PHOSPHO_SITE (pattern) Casein kinase II phosphorylation site	209-212; 225-228; 287-290; 301-304; 346-349; 493-496; 506-509; 553-556; 597-600; 623-626; 722-725; 836-839; 871-874; 886-889; 986-989; 1010-1013; 1046-1049; 1085-1088; 1093-1096; 1193-1196; 1254-1257
PS00007 TYR_PHOSPHO_SITE (pattern) Tyrosine kinase phosphorylation site	739-746
PS00004 CAMP_PHOSPHO_SITE (pattern) cAMP- and cGMP-dependent protein kinase phosphorylation site	993-996; 1070-1073
PS00003 SULFATION (rule) Tyrosine sulfation site	823-837

[0185] *ClustalW Polypeptide Alignment and Sequence Analysis to Estimate the Potential Effect of FLT-4 mutation on FLT-4 Function.* ClustalW polypeptide alignment and sequence analysis was used to estimate the effect of FLT-4 mutation on FLT-4 biological function.

Known FLT-4 sequences of various organisms including, *e.g.*, mouse KDR; rat KDR; human KDR; chicken KDR; rat_2 FLT4; rat FLT4; mouse FLT4; human FLT4; chicken VEGFR3; mouse FLT1; rat FLT1; human FLT1; chicken FLT1; chicken VEGFR2; and fly KDR sequences were obtained from GenBank and aligned using ClustalW. Chenna *et al.*, *Nucleic Acids Res.*, 31 (13):3497-500 (2003). ClustalW is a general purpose multiple sequence alignment program for DNA or proteins. It produces biologically meaningful multiple sequence alignments of divergent sequences. It calculates the best match for the selected sequences, and lines them up so that the identities, similarities and differences can be seen.

-59-

[0186] For the position with the mutation reported, the mutated residue was inspected for its occurrence in organisms other than human. It was hypothesized that if the mutated residue was present in the wild-type sequence of another species in the corresponding position, the amino acid change may not have any adverse effect on the protein function. The results of the Clustal W comparison analysis are summarized below in TABLE 19 and TABLE 20. The mutated amino acid residues identified in the present studies is highlighted in bold, underlined text.

TABLE 19
Summary of Sequence Alignment of FLT-4 Sequences from Multiple Organisms

KDR_mouse	-----
KDR_rat	-----
KDR_human	-----
KDR_chicken	-----
FLT4_rat_2	-----
FLT4_rat	-----
FLT4_mouse	-----
FLT4_human	-----
VEGFR3_chicken	-----
FLT1_mouse	-----
FLT1_rat	-----
FLT1_human	-----
FLT1_chicken	-----
VEGFR2_chicken	MSRFLPRLSLGVEGESTPREKEPGAFLQSRALSPAHRWFTEIRQHTHQCVTAAQELRF
KDR_fly	-----
KDR_mouse	-----
KDR_rat	-----
KDR_human	-----
KDR_chicken	-----
FLT4_rat_2	-----
FLT4_rat	-----
FLT4_mouse	-----
FLT4_human	-----
VEGFR3_chicken	-----
FLT1_mouse	-----
FLT1_rat	-----
FLT1_human	-----
FLT1_chicken	-----
VEGFR2_chicken	QEQLHCLPALSVLQVPAWSIPSTHTPIFEKYVSIYRGQNVVYQNSAPLSKFSNDREGTGTG
KDR_fly	-----
KDR_mouse	-----MESKALLA-----
KDR_rat	-----MESRALLA-----
KDR_human	-----MQSKVLLA-----
KDR_chicken	-----MELGFLRV-----
FLT4_rat_2	-----MQPGAALN-----
FLT4_rat	-----MQPGAALN-----
FLT4_mouse	-----MQPGAALN-----
FLT4_human	-----MQPGAALN-----
VEGFR3_chicken	-----MTVRSTVLMENSESNRYFEMFVHE-----
FLT1_mouse	-----MVSCWDTA-----
FLT1_rat	-----MVSCWDTA-----
FLT1_human	-----MVSYWDTG-----
FLT1_chicken	-----MPRQLLSG-----
VEGFR2_chicken	SLQSHPVSLNCCLEASLSASGGKGDCKGDTKGTARQPIAAQKEHQDRKAF-----
KDR_fly	-----MTPLKTSVKAF-----

-60-

```

KDR_mouse      --VALWFCVETRAAS-----VGLPGDFLHPPKLST
KDR_rat         --VALWFCVETRAAS-----VGLPGDSLHPPKLST
KDR_human       --VALWLCVETRAAS-----VGLPSVSLDLPLSLI
KDR_chicken     --LTVLLCLPPVFTG-----LFIS---MDQPTLSI
FLT4_rat_2      --LRLWLCLGLLQG-----LANGYSMTPTLNI
FLT4_rat        --RRLWLCLGLLQG-----LANGYSMTPTLNI
FLT4_mouse      --LRLWLCLGLLQG-----LANGYSMTPTLNI
FLT4_human      --LRLWLCLGLLDG-----LVSGYSMTPTLNI
VEGFR3_chicken  QLARCVSCTACRRY-----LVSSYSMTPTLSI
FLT1_mouse      --VLPYALLGCLLLT-----GYSGSKLKVPELSL
FLT1_rat        --VLPYALLGCLLLT-----GYSGSKLKVPELSL
FLT1_human      --VLLCALLSCLLLT-----GSSSGSKLKVPELSL
FLT1_chicken    --TVLLG--AAFLLA-----GSTSGSKLKVPELSV
VEGFR2_chicken  --ADTDVAIAALAEHTGNRTPREEGRAVRRSGEWRRFNRSQQVGAVRASRSPAPARS
KDR_fly         --FTLHVLFESCISH-----GLVEGSRLPDPQLLP

```

```

KDR_mouse      QK-----DILTILANT
KDR_rat        QK-----DILTILANT
KDR_human      QK-----DILTIKANT
KDR_chicken    QK-----SVLTITTND
FLT4_rat_2     TE-----DSYVIDTGD
FLT4_rat       TE-----DSYVIDTGD
FLT4_mouse     TE-----DSYVIDTGD
FLT4_human     TE-----ESHVIDTGD
VEGFR3_chicken TE-----EEHVINAkd
FLT1_mouse     KG-----TQHVMQAGQ
FLT1_rat       KG-----TQHVMQAGQ
FLT1_human     KG-----TQHIMQAGQ
FLT1_chicken   NG-----RQHVVQAGQ
VEGFR2_chicken DGPKETSPPRAAPAMLGGPSAALACALQALLTLGLAEDEFFKPVLSIAKDHIIVLPGE
KDR_fly        DG-----DTHLQHVGE

```

```

KDR_mouse      TLQITCR-----GQRD-----LDWLWPNAQ-----R
KDR_rat        TLQITCR-----GQRD-----LDWLWPNTF-----R
KDR_human      TLQITCR-----GQRD-----LDWLWPNNQ-----S
KDR_chicken    TLNITCS-----GQRA-----VYWSWPNQ-----S
FLT4_rat_2     SLSISCR-----GQHP-----LEWTWPGAQEVLTGGK
FLT4_rat       SLSISCR-----GQHP-----LEWTWPGAQEVLTGGK
FLT4_mouse     SLSISCR-----GQHP-----LEWTWPGAQEVLTGGK
FLT4_human     SLSISCR-----GQHP-----LEWAWPGAQEPATGDK
VEGFR3_chicken TLTITCRTGDLCSIPPDTDHARAGWKPCRCFDRGMWHSRVKGLVDYSGRKEVDFDGPAL
FLT1_mouse     TLFLKCR-----GEAA-----HSWSLPTTV-----S
FLT1_rat       TLFLKCR-----GEAA-----HSWSLPTTV-----S
FLT1_human     TLHLQCR-----GEAA-----HKWSLPEMV-----S
FLT1_chicken   TLNLTCR-----GEML-----HSWSLPEAL-----S
VEGFR2_chicken DLKIKCR-----GARP-----VSWLLGEAQ-----R
KDR_fly        TLTLICR-----GSTA-----LHWRL--AS-----R

```

```

KDR_mouse      DS--EERVLVTECGG-GDSIFCKTLTIPRVVGNNDTGAYKCSYR-----DVDIASTVYVY
KDR_rat        DS--EERVLVTECG--DSIFCKTLTIPRVVGNNDTGAYKCFYR-----DIDVSSIVYVY
KDR_human      GS--EQRVEVTECS---DGLFCKTLTIPKVIGNDTGAYKCFYR-----ETDLASVIYVY
KDR_chicken    NA--EKRLAVTGCS---DGPFCRKLTLRLVIGNDTGDYRCLYG-----DSQAATTIYVY
FLT4_rat_2     DS--EDTQVVQDCEGTEARPYCKVLSLAQTHANNTGSYYCYKYIKARIEGTTAASYVF
FLT4_rat       DS--EDTQVVQDCEGTEARPYCKVLSLAQTHANNTGSYYCYKYIKARIEGTTAASYVF
FLT4_mouse     DS--EDTRVVHDCEGTEARPYCKVLLLAQTHANNTGSYHCYKYIKARIEGTTAASYVF
FLT4_human     DS--EDTGVRDCEGTDARPYCKVLLLEHVHNDTGSYVCYKYIKARIEGTTAASSYVF
VEGFR3_chicken LAPAEEMSAVTGMEEPEDQ-AVPCMAPQWCLGTYHTCCSFLRVCDWCWVLPCCPPTAYS
FLT1_mouse     QEDKRLSITPPSACGRDNRFQFCSTLTLDTAQANHTGLYTCRYLPTSTSKKKKAESSYIF
FLT1_rat       QEDKRLSVTR-SACGRNRRQFCSTLTLNMAQANHTGLYSCRYLPKSTSKKKMESAIYIF
FLT1_human     KESERLSITK-SACGRNGKQFCSTLTLNMAQANHTGFYSCRYLPKSTSKKKKETESAIYIF
FLT1_chicken   KDSKRLNVTK-YACGRNGTQFCSTLTLSRTQANDTGRYSCRYLTPVKKKR--ESIVYVF
VEGFR2_chicken KDP-RVDISS---CRNNSRQSCSTLLRNVTASDTGYLTCVHDDAPS--REELTAKIYVF
KDR_fly        NVS-SVRIES---CEERLHKHCSKLVHNLRHNDTGIYSCSHKKSSD--HEVST---YVF

```

-61-

```

KDR_mouse      VR-----DYSRSPFIASVSDQHGIYITENKNKTVVIPCRGSISNLNVSLCARYPE
KDR_rat        VQ-----DHRSPFIASVSDQHGIVYITENKNKTVVIPCRGSISNLNVSLCARYPE
KDR_human      VQ-----DYSRSPFIASVSDQHGVYITENKNKTVVIPCLGSISNLNVSLCARYPE
KDR_chicken    VQ-----DYSRSPFVTSVGDQLGIVYIT--KNKTVVVPCLGTVSNLNVSLHAKYPE
FLT4_rat_2     VR-----DFEQPFINKP-----DTLLVNRKDSMWVPCLVSIPLNITLRSQSS-
FLT4_rat       VR-----DFEQPFINKP-----DTLLVNRKDSMWVPCLVSIPLNITLRSQSS-
FLT4_mouse     VR-----DFKHPPINKP-----DTLLVNRKDSMWVPCLVSIPLNITLRSQSS-
FLT4_human     VR-----DFEQPFINKP-----DTLLVNRKDSMWVPCLVSIPLNITLRSQSS-
VEGFR3_chicken IQQDKRVLSPGSNFEQPFINKP-----ETLLISKENTWVPCLVSIPLNITLRSQSS-
FLT1_mouse     VS-----DAGSPFIEMHTDIPKLVHMTTEGR--QLIIPCRVTSPNVTVTLK-KFPF
FLT1_rat       VS-----DAGSPFIEMHSDIPKLVHMTTEGR--ELIIPCRVTSPNITVTLK-KFPF
FLT1_human     IS-----DTGRPFVEMYSEIPEIHMTEGR--ELVIPCRVTSPNITVTLK-KFPF
FLT1_chicken   IN-----DTSNPFVEMHSDIPKIIHMTVGK--EMIIPCRVTAPNIAVTLK-KIPR
VEGFR2_chicken VE-----DYRNPFVEFYEPHQIIYLLEAKK--TVVIPCRVTSPDIRPQLM-QYPH
KDR_fly        VK-----DPHHPFVEAY-SLPHPLFAYRNDP-YFVVPCTRTPNQNVILE-TQMN
               :           :   ** :           : ** : ..   *

```

```

KDR_mouse      KRFVPDGNRISWDSEIGFTLPSYMSIYAGMVCFEAKINDETYQSIMYIVVVVGYRIYDVI
KDR_rat        KRFVPDGNRISWDSEKGFITPSYMSIYAGMVCFEAKINDETYQSIMYIVLVVGYRIYDVV
KDR_human      KRFVPDGNRISWDSKKGFTIPSYMSIYAGMVCFEAKINDESYQSIMYIVVVVGYRIYDVV
KDR_chicken    KVFVPDGSISWDNKKGFTIPSHLISYAGMVCFEAKIDNESYQSVIYIVAVVGYRIYDLT
FLT4_rat_2     -VLHPDGGQEVWLDDRRGMVPTLLLRDALYLCETTWDGQDQFLSNPFLVHITGNELYDIQ
FLT4_rat       -VLHPDGGQEVWLDDRRGMVPTLLLRDALYLCETTWDGQDQFLSNPFLVHITGNELYDIQ
FLT4_mouse     -ALHPDGGQEVWLDDRRGMVPTQLLRDALYLCETTWDGQDQFLSNLFFVHITGNELYDIQ
FLT4_human     -VLHPDGGQEVWLDDRRGMVPTLLLRDALYLCETTWDGQDQFLSNPFLVHITGNELYDIQ
VEGFR3_chicken -LIHPDRKAI FWDNKKGMQVPTQLIRDSLFVQCETVIDKKVFKSNFFIIHAGIELYDIQ
FLT1_mouse     DTLPDGGQRIWDSRRGFI IANATYKEIGLLNCEATVNGHLYQT-NYLTHRQTNTILDVQ
FLT1_rat       DALTPDGGQRIAWDSRRGFI IANATYKEIGLLTCEATVNGHLYQT-SYLTHRQTNTILDVQ
FLT1_human     DTLPDGGKRIIWDNRKGFII SNATYKEIGLLTCEATVNGHLYKT-NYLTHRQTNTIIDVQ
FLT1_chicken   ETLPDGGKTI IWDNRMGFRIP EATYRFIGLLSCETTIGGHKYST-KYLTHRENTNIFDIK
VEGFR2_chicken REKLKFKK-MVWDPKRGFVIPS AFTYSGVLTCTAGINGSVFTS-YVLQKLEDRVQNLA
KDR_fly        PMADDVKRGVQWDPKKGFTVPLKPYDSYHLITCLTRVDNAEFSS-VYLLKRLTMEIKNLA
               . : ** * : ..           : * : . : : : : : : :

```

```

KDR_mouse      LSPPEIELSAGEKLVNLCTARTELNVGLDFTWHSPPSKSHKKIVNRDVKPFPGTVAKM
KDR_rat        LSPPEIELSAGEKLVNLCTARTELNVGLDFSWQFPSSKHQHKKIVNRDVKSLPGTVAKM
KDR_human      LSPSHGIELSVGEKLVNLCTARTELNVGIDFNWEYPPSKHQHKKLVNRDLKTQSGSEMKK
KDR_chicken    MNPHYQVELAVGEKLVNLCTVRTELNVGIDFKWDYPSIK--EKRAITRDLKTTAG-EIKT
FLT4_rat_2     LYPKKSLELLVGEKLVNLCTVWAEFDSGVTFDWDYPGKQAEKAKWVP---ERRSQQTHTF
FLT4_rat       LYPKKSLELLVGEKLVNLCTVWAEFDSGVTFDWDYPGKQAEKAKWVP---ERRSQQTHTF
FLT4_mouse     LYPKKSLELLVGEKLVNLCTVWAEFDSGVTFDWDYPGKQAEKAKWVP---ERRSQQTHTF
FLT4_human     LLPRKSLELLVGEKLVNLCTVWAEFNSGVTFDWDYPGKQAEKAKWVP---ERRSQQTHTF
VEGFR3_chicken LYPKKAMELLVGEKLVNLCTVWAEFNSGVRFQWTPYPGKQMAKRAVIES---ERRSLQTHTE
FLT1_mouse     IRPSPVRLHGGQTLVNLCTATTELNTRVQMSWNPYPGKATKRASIRQRI--DRSHSHNNV
FLT1_rat       ISPPSPVRFLRGQTLVNLCTVTTDLNTRVQMSWNPYPGKATKRASIRQRI--DQSNPHSNV
FLT1_human     ISTPRPVKLLRGHTLVNLCTATTPLNTRVQMTWSYPDEKNKRASVRRRI--DQSNSHANI
FLT1_chicken   LSTPRLVKLLKGDSLAINCTVKAAWNTRVQMTWTYPGEAMKRGSVTQRI--DQKNREANV
VEGFR2_chicken LK-ATPKKLLVGETLYLECKAETFINGRIFIEWTCPNGKHPYPRGMD----RSKPVYK
KDR_fly        IE-PERPRVLVGDTLILNCSAETTYNGRIHFWEFFHKERINRPHHFSTT--PVQLAQIMV
               :           .. *..* : : : : : : : :

```

```

KDR_mouse      FLSTLTIESVTKSDQGEYTCVASSGRMIKRNRTFVRVHTKPFIAFGSGMKSLVEATVG-S
KDR_rat        FLSTLTIDSVTKSDQGEYTCAYSGLMTKKNKTFFVRVHTKPFIAFGSGMKSLVEATVG-S
KDR_human      FLSTLTIDGVTRSDQGLYTCAASSGLMTKKNSTFVRVHEKPFVAFSGMESLVEATVG-E
KDR_chicken    FVSTLTIDSVNLSDKGRYTCAASSGRMNKNSYFIIHESPFIHLEK-MENVVEMKLG-D
FLT4_rat_2     LSSILTIHNVSQHDLGPVCEANNGIQRFRESTEVIHVEKPFISVEWLKGPVLEATAGDE
FLT4_rat       LSSILTIHNVSQHDLGPVCEANNGIQRFRESTEVIHVEKPFISVEWLKGPVLEATAGDE
FLT4_mouse     LSSILTIHNVSQHDLGPVCEANNGIQRFRESTEVIHVEKPFISVEWLKGPVLEATAGDE
FLT4_human     LSSILTIHNVSQHDLGSYVCKANNGIQRFRESTEVIHVENPFISVEWLKGPVLEATAGDE
VEGFR3_chicken LSSILTIHNVSQDLGRYTCTATNGAQMLEESTDVIVHEKPFINVEWRKGPVLEATAGDE
FLT1_mouse     FHSVLKINNVSERDKGLYTCRVKSGSSSFNTSVHVYKGFISVKHRKQPVQETTAGRR
FLT1_rat       FHSVLKINNVSERDKGLYTCRVKSGSSSFNTSVHVYKGFISVKHRKQPVQETIAGKR
FLT1_human     FYSVLTIIDKMQNKDKGLYTCRVKSGSSSFNTSVHIDYKAFITVKHRKQPVLETVAGKR
FLT1_chicken   FYSILVIDKVRIDKGOYACHVKSGSPSNKLVNTTVIVYDKRFINLKRKRKTMLEAVAGRK
VEGFR2_chicken FGSSLTIENTVMTEDGGLYICR-TFNLTSLANVTVTYVDKPFLEVSHKGLYIETSGHK
KDR_fly        MSKALIVPNVTMEDKGTCTGSIIEFKLQMSKTVIVYEHFPLNVTHNKRKFTSTVEGRR
               : . * : : * * * * . : : * : . *

```

-62-

```

KDR_mouse      QVRIPVKYLSYPAPD-IKWYRNGRPIESNYTMIVG----DELTIMEVTERDAGNYTVILT
KDR_rat        QVRIPVKYLSYPAPD-IKWYRNGRPIESNYTMIVG----DELTIMEVSESDAGNYTVILT
KDR_human      RVRIPAKYLGYPPE-IKWYKNGIPLESNHTIKAG----HVLTIMEVSESDTGNVTVILT
KDR_chicken    TVSI PVKFGYPSPE-AKWYKNGKVMNANHTVKLG----YTLVITDATEKDAGNYTVVLT
FLT4_rat_2     MVKLPVKLAAYPPE-FQWYKDRKAVTGRHNPHA-----LVLKEVTEASAGVYTLALW
FLT4_rat       MVKLPVKLAAYPPE-FQWYKDRKAVTGRHNPHA-----LVLKEVTEASAGVYTLALW
FLT4_mouse     LVKLPVKLAAYPPE-FQWYKDRKAVTGRHNPHA-----LVLKEVTEASAGVYTLALW
FLT4_human     LVKLPVKLAAYPPE-FQWYKDGKALSGRHSpha-----LVLKEVTEASTGTYYTLALW
VEGFR3_chicken AVKLPVKVAYPPPD-FQWYKDGKLIP-KQSQSS-----MQIKDVAEHHAGTYTLVLR
FLT1_mouse     SYRLSMKVKAFPSPE-IVWLKDGSPATLKSARYLV--HGYSLLIKDVTEDAGDYTILLG
FLT1_rat       SHRLSMKVKAFPSPE-VVWLKDGVPATEKSARYSV--HGYSLLIKDVTEDAGDYTILLG
FLT1_human     SYRLSMKVKAFPSPE-VVWLKDGVPATEKSARYLT--RGYSLLIKDVTEDAGNYTILLS
FLT1_chicken   SYRLPMKVKAFPSPE-VTWLKDGLPAAEKARYMV--KNYSLLIKDVAEEDAGNYTILLS
VEGFR2_chicken SLKLAAKVDAFPRPT-VTWYKDGKLIKDNQSCYEPDPAKYRLGIRDVRPKHAGNYTIVLS
KDR_fly        -VQFEPRVNAVAPDRVLWYKDGVAISENSTCYET--AGYNLTIKQVRQKDAGITIALS
               : : . * * * : : : : : * : * : *

KDR_mouse      NPISMEKQSHMVSLVVNVPPQIGEKALI--SPMDSYQYGTMTLTCTVYANPPLHHIQWY
KDR_rat        NPISMEKQSHMVSLVVNVPPQIGEKALI--SPMDSYQYGTMTLTCTVYANPPLHHIQWY
KDR_human      NPISKEKQSHVSLVVVPPQIGEXSLI--SPVDSYQYGTMTLTCTVYAI PPPHHIHWY
KDR_chicken    NPTNKMQRHTFTLLVNVPPQIGENALM--APVDSYKYGSTQTLTCTIYAVPPAAVLWY
FLT4_rat_2     NSAAGLRQNISLELVNVVPPPIHEKEAS--SPS-IYSRHSRQTLTCTTYGVPQPLSVQWH
FLT4_rat       NSAAGLRQNISLELVNVVPPPIHEKEAS--SPS-IYSRHSRQTLTCTTYGVPQPLSVQWH
FLT4_mouse     NSAAGLRQNISLELVNVVPPPIHEKEAS--SPS-IYSRHSRQTLTCTAYGVPQPLSVQWH
FLT4_human     NSAAGLRQNISLELVNVVPPPIHEKEAS--SPS-IYSRHSRQTLTCTAYGVPPLPLSIQWH
VEGFR3_chicken NRLVGLEKHISLQLIVNVPPRIHEKETS--SPS-IYSRRSPQALTCTVYGIPAPEVIQWQ
FLT1_mouse     IKQSLRFLKNTATLIVNVKPKIYEKSVSSLPSPPLYPLGSRQVLTCTVYGI PRPT-ITWL
FLT1_rat       IKQSKLFRNLATLIVNVKPKIYEKSVSSLPSPPLYPLGSRQVLTCTVYGI PQPT-IKWL
FLT1_human     IKQSNVEKNLTATLIVNVKPKIYEKAVSSFPDPALYPLGSRQILTCTAYGI PQPT-IKWF
FLT1_chicken   LRQWNLSKNLTVTLVNVKPKIYENAVSSFPDPNLYLLSSKQVLTCTVYGI PPPK-ITWM
VEGFR2_chicken NAKHGLYKNLTMLQVLVTERPKIYEKETD-FDKIERVELRSEHKIQCTVSGNPDPK-VEWK
KDR_fly        NQERGLYRNISYKLEVRVKPKIFEEDVA-PAGPQTFRYDQRHKLTCTAFGI PMPN-ITWF
               : * * * : * : : : * : * : *

KDR_mouse      WQLEEACSYRPGQ-----TSPYACKEWRRHVEDFQGGNKIEVTKNQYALIEG
KDR_rat        WQLEEACSYRPSQ-----TNPYTCKEWRHVKDFQGGNKIEVTKNQYALIEG
KDR_human      WQLEEACANEPSQAVS-----VTNPYPCEEWRSVEDFQGGNKIEVKNQYALIEG
KDR_chicken    WQLEEECTFSPQKVRL-----GANPYACRKKKVISERKGGNQVEIKQR-VVTIAG
FLT4_rat_2     WRPWT PCKTFAQRSLR-----RRQPRDGMPQCRDWKEVTTQDAVNPIESLDTWTESVEG
FLT4_rat       WRPWT PCKTFAQRSLR-----RRQPRDGMPQCRDWKEVTTQDAVNPIESLDTWTESVEG
FLT4_mouse     WRPWT PCKTFAQRSLR-----RRQPRDGMPQCRDWKEVTTQDAVNPIESLDSWTEFVEG
FLT4_human     WRPWT PCKMFAQRSLR-----RRQQQDLMQCRDWRAVTTQDAVNPIESLDTWTESFVEG
VEGFR3_chicken WRPWMPCRMFSRRSLNSRHRRAARRHQDRMPECKDWKDVSSQDAVNPIESIDTWEFVEG
FLT1_mouse     WHPCHHNHHSKERYDFCT-----ENEESFILDPSNLGNRIESISQRMVIEG
FLT1_rat       WHPCHYHNHHSKERNDFCF-----GSEESFILDSSSNIGNRIEGITQRMVIEG
FLT1_human     WHPCHNHNHSEARCDPCS-----NNEESFILDADSNMGNRIESITQRMVIEG
FLT1_chicken   WYPCRQNHHSKTRRGFCS-----RTDGSFNLKTSNIGNRIQSI IERTAIIEG
VEGFR2_chicken WQPCSLTD--TRCNPK-----GQKIVVQ-ENMFIGNKIKSIENITMKHGD
KDR_fly        WQPCDPSANLTECKLY-----TDPLPIENVDDFPQNPDKVNSKVGLLKS
               * * : :

KDR_mouse      KNK-----TVST
KDR_rat        KNK-----TVST
KDR_human      KNK-----TVST
KDR_chicken    KTK-----TVST
FLT4_rat_2     KNK-----TVSK
FLT4_rat       KNK-----TVSK
FLT4_mouse     KNK-----TVSK
FLT4_human     KNK-----TVSK
VEGFR3_chicken RNKALPVPRHRQGFVTLPRGWQGMQLLREOPKGEAGICFLSLSSFPVFSYRLPLTVSK
FLT1_mouse     TNK-----TVST
FLT1_rat       TNK-----TVST
FLT1_human     KNK-----MAST
FLT1_chicken   KNK-----TAST
VEGFR2_chicken KRK-----IVST
KDR_fly        KNR-----TIST
               : *

```

-63-

```

KDR_mouse      LVIQAANVSALYKCEAINKAGRGERSVSHFVIRGPE---ITVQPAAQPTQESVSLLCTA
KDR_rat        LVIQAAYVSALYKCEAINKAGRGERSVSHFVIRGPE---ITVQPATQPTERESMSLLCTA
KDR_human      LVIQAANVSALYKCEAVNKVGRGERVISFHVTRGPE---ITLQPDQMPTQESVSLLWCTA
KDR_chicken    LVIQAANVSALYRCMATNRAGSSSERVISFHVTRGLE---INLQPRSQTLEKDNSTLSQCTA
FLT4_rat_2     LVIQDANVSAMKCVVFNKVQDERLIYFYVTTIPDGFSIESEPSDPLEGQSVRLSCRA
FLT4_rat       LVIQDANVSAMKCVVFNKVQDERLIYFYVTTIPDGFSIESEPSDPLEGQSVRLSCRA
FLT4_mouse     LVIQDANVSAMKCVVFNKVQDERLIYFYVTTIPDGFSIESEPSDPLEGQSVRLSCRA
FLT4_human     LVIQANVSAMKCVVSNKVQDERLIYFYVTTIPDGFTIESKPSEELLEGGQPVLLSCQA
VEGFR3_chicken LAIQEANVSAMKCIASNKVGRDERLIYFYVTTIPDGFEIESQPSEEPLEGQDLQLSCNA
FLT1_mouse     LVVADSQTPIYSCRAFNIKIGTVERNIKFYVTDVPNG--FHVSLKMPAEGEDLKLSCVV
FLT1_rat       LVVADSRTPGSYSCAFNIKIGTVRDIRFYVTDVPNG--FHVSLKIPTEGEDLKLSCVV
FLT1_human     LVVADSRSIGIYICIASNKVGTVGRNISFYITDVPNG--FHVNLKMPTEGEDLKLSCVT
FLT1_chicken   LVVAEAKSSGIYSCVASNKVGAERNVSVFLVTDVPNG--FHISLEKVPTEGENLVLSCSA
VEGFR2_chicken LTMENSSTSGIYCVASNKVGEEERSIEFYVSDVPNG--LQTAPQVTAIAGSTAQLTCRA
KDR_fly        LVVKTANVSGVYSCTARNELGNRTMRIPFYVDDHPQP--FEIEP-STAVAGDDITLTCTRG
*.: : .. * * . * . : * : . : . * *

KDR_mouse      DRNTFENLTWYKLGSAQTSVHMGESLTPVCKNLDALWK---LNGTMFS-NSTNDILIVAF
KDR_rat        DRNTFENLTWYKLGSAQTSVHMGESLTPVCKNLDALWK---LNGTVFS-NSTNDILIVAF
KDR_human      DRSTFENLTWYKLGPOPLPIHVGEPLTPVCKNLDLWK---LNATMFS-NSTNDILIMEL
KDR_chicken    DKFTFEKLSWYKLSAHALQTPFGGLPMPVCKNLDALQK---LNATVSNNTNGENLTLELIL
FLT4_rat_2     DNYTYEHLRWYRLNLSLTHDAQGNPLLLDCKNVHLFATPLEANLEEAEPGARHATLSLNI
FLT4_rat       DNYTYEHLRWYRLNLSLTHDAQGNPLLLDCKNVHLFATPLEANLEEAEPGARHATLSLNI
FLT4_mouse     DNYTYEHLRWYRLNLSLTHDAQGNPLLLDCKNVHLFATPLEANLEEAEPGARHATLSLNI
FLT4_human     DSYKYEHLRWYRLNLSLTHDAHGNPLLLDCKNVHLFATPLAASLEEVAPGARHATLSLSLI
VEGFR3_chicken DNYTYENLQWYRLNLSKLHDEEGNPLVLDCKNVHHYATKMQGELR-FQPSNDATLLTLTI
FLT1_mouse     NKFLYRDITWILLRTVNNTMHS-----ISKQKMATQDYSITLNLVI
FLT1_rat       SKFLYRDITWILLRTVNNTMHS-----ISKQKMATQDYSITLNLVI
FLT1_human     NKFLYRDVTWILLRTVNNTMHS-----ISKQKMAITKEHSITLNLTI
FLT1_chicken   NKFMKYKDISWILPRTVTN-----QTKARKALNKEYSITLTLTI
VEGFR2_chicken SRVIYSNLAWSYPSLGAH-----GSSLTRSADAYSISLTLII
KDR_fly        TRYLYDRLTWYDPLGYKVP-----KDETLRIEPTYTISLSIKL
: : * * . * * : * : . : . * *

KDR_mouse      QNASLQDQGDYVCSAQDKKTK-KRHCLVKQLIILERMAMITGNLENQTTTIGETIEVTC
KDR_rat        QNASLQDQGDYVCSAQDKKTK-KRHCLVKQLVILERMAMITGNLENQTTTIGETIEVVC
KDR_human      KNASLQDQGDYVCLAQDRKTK-KRHCVVRQLTVLERVAPTITGNLENQTTSIGESIEVSC
KDR_chicken    RNISLQDGGDYVCSAQDKKAK-TQHCLVKHLTVQEPMAPTLVGNLENQTTNIGETIEVSC
FLT4_rat_2     PRVAPEDGDDYVCEVQDRRNQ-DKHCHKKYLSVQALEAPRLTQNLTDLLVNVSDSLEMRC
FLT4_rat       PRVAPEDGDDYVCEVQDRRSQ-DKHCHKKYLSVQALEAPRLTQNLTDLLVNVRTSLEMRC
FLT4_mouse     PRVAPEDGDDYVCEVQDRRSQ-DKHCHKKYLSVQALEAPRLTQNLTDLLVNVSDSLEMRC
FLT4_human     PRVAPHEGHYVCEVQDRRSQ-DKHCHKKYLSVQALEAPRLTQNLTDLLVNVSDSLEMRC
VEGFR3_chicken PNISLGEEDDYVCEVQNRKTR-EKHCHKKYISVQALEIPRLKQNLTDIWNVSDSIEMRC
FLT1_mouse     KNVSLDSGTYACRARNIYTG-EDILRKTEVLVRDSEAPHLQLNLSDEYSISGSTTLDC
FLT1_rat       KNVSLDSGTYACRARNIYTG-EEILRKTEVLVRDLEAPHLQLNLSDEYSISGSTTLDC
FLT1_human     MNVSLQDSGTYACRARNVYTG-EEILQKKEITIRDQEAAPYLLRNLSDHTVAISSSTTLDC
FLT1_chicken   RNVSLAHSGTYTCRARNIFTG-KEVLQKKDVSIQAQEPALLRQLMDQTVNTSNSAMLEC
VEGFR2_chicken ANVTEHSGLYKCAQHQHNGSDTLEQHTRLIRAKAAPYVLQNLTDLEVNISGKIILEC
KDR_fly        PNVSRNHTLGYECQALNINTN-KVVNVTSALTIDERQGPWLMQNLTNQDVNSSSTLTILAC
. : . * * . . : : * : * : . : . *

KDR_mouse      PASGNPTPHITWFKDNETLVEDSGIVLRDGNRNLTIIRVRKEDGGLYTCQACNVLGCARA
KDR_rat        PTSGNPTPLITWFKDNETLVEDSGIVLRDGNRNLTIIRVRKEDGGLYTCQACNVLGCARA
KDR_human      TASGNPPPQIMWFKDNETLVEDSGIVLRDGNRNLTIIRVRKEDGGLYTCQACSVLGCAKV
KDR_chicken    TVNGIPPPNITWFKNGETLFEDSGIVLRDGNKTLTIIRVRKEDGGLYTCLACNIGCKKA
FLT4_rat_2     PVAGAHVPSIVWYKDERLLEKESGIDLADSNQRLSIQVRREEDAGRYLCSVCNAKGCVNS
FLT4_rat       PVAGAHVPSIVWYKDERLLEKESGIDLADSNQRLSIQVRREEDAGRYLCSVCNAKGCVNS
FLT4_mouse     PVAGAHVPSIVWYKDERLLEKESGIDLADSNQRLSIQVRREEDAGRYLCSVCNAKGCVNS
FLT4_human     LVAGAHAPSIVWYKDERLLEKESGVDLADSNQRLSIQVRREEDAGRYLCSVCNAKGCVNS
VEGFR3_chicken KVDGNHVPDISWYKDEKLVEEVSIGIDLADFNQRLSIQVRREEDAGLYLCSVCNAKGCVNS
FLT1_mouse     QARGVPAPQITWFKNNHKIQEPGIIILGPGNSTLFIERVTEDEGVYRCRATNQKGVES
FLT1_rat       QARGVPAPQITWFKNNHKIQEPGIIILGPGNSTLFIERVTEDEGVYRCRATNQKGVES
FLT1_human     HANGVPEPQITWFKNNHKIQEPGIIILGPGNSTLFIERVTEDEGVYHCKATNQKGSVES
FLT1_chicken   QVHGIPPEPQITWFKNHEEIQESGIIILGPGSRMLFIERVKEEDEGLYQCIATNLKGSVES
VEGFR2_chicken RVSGTPEPQVTSKDGYPISAASGISMENN--TLVIERVKKDEGLYECKATNMQDQST
KDR_fly        LAYGVPAPFITWYKDKTPVTEGPGITLKDGG-TLIIERVKKDEGIYECRASNDGGEAKT
. * * : * * : . * : : * * . * : * * * . . *

```



```

KDR_mouse      ICDFGLARDIYKDPDYVRKGDARLPLKWMAPETIFDRVYTIQSDVWSFGVLLWEIFSLGA
KDR_rat        ICDFGLARDIYKDPDYVRKGDPRPLPLKWMAPETIFDRIYTIQSGVWSFGVLLWEIFSLGA
KDR_human      ICDFGLARDIYKDPDYVRKGDARLPLKWMAPETIFDRVYTIQSDVWSFGVLLWEIFSLGA
KDR_chicken    ICDFGLARDIYKDPDYVRKGDARLPLKWMAPETIFDRVYTIQSDVWSFGVLLWEIFSLGA
FLT4_rat_2     ICDFGLARDIYKDPDYVRKGSARLPLKWMAPESIFDKVYTTQSDVWSFGVLLWEIFSLGA
FLT4_rat       ICDFGLARDIYKDPDYVRKGSARLPLKWMAPESIFDKVYTTQSDVWSFGVLLWEIFSLGA
FLT4_mouse     ICDFGLARDIYKDPDYVRKGSARLPLKWMAPESIFDKVYTTQSDVWSFGVLLWEIFSLGA
FLT4_human     ICDFGLARDIYKDPDYVRKGSARLPLKWMAPESIFDKVYTTQSDVWSFGVLLWEIFSLGA
VEGFR3_chicken ICDFGLARDIYKDPDYVRKGSARLPLKWMAPESIFDKVYTTQSDVWSFGVLLWEIFSLGA
FLT1_mouse     ICDFGLARDIYKNPDYVRRGDTRLPLKWMAPESIFDKVYSTKSDVWSYGVLLWEIFSLGG
FLT1_rat       ICDFGLARDIYKNPDYVRRGDTRLPLKWMAPESIFDKVYSTKSDVWSYGVLLWEIFSLGG
FLT1_human     ICDFGLARDIYKNPDYVRKGDTRLPLKWMAPESIFDKIYSTKSDVWSYGVLLWEIFSLGG
FLT1_chicken   ICDFGLARDIYKNPDYVRKGDARLPLKWMAPESIFDKIYNTKSDVWSYGVLLWEIFSLGA
VEGFR2_chicken ICDFGLARDIYKDPDYVRKGDARLPLKWMAPAEAFDKIYTTQSDVWSFGVLLWEIFSLGA
KDR_fly        ICDFGLARDIYKDPDYVRKGDARLPLKWMAPAEAFDKIYTTQSDVWSFGVLMWEIFSLGA
*****

```

```
KDR_mouse SPYPGVKIDEEFCRRLKEGTRMRAPDYTTPEMYQTMLDCWHEDPNQRFSELSVEHLGNL
KDR_rat SPYPGVKIDDKFCRRLKEGTRMRAPDYTTPEMYQTMLDCWHEDPNQRPAFSELVEHLGNL
KDR_human SPYPGVKIDEEFCRRLKEGTRMRAPDYTTPEMYQTMLDCWGHGPSQRTFSELVEHLGNL
KDR_chicken SPYPGVKIDEEFCRRLKEGTRMRAPDYTTPEMYQTMLDCWHGDPKQRTFSELVEHLGNL
FLT4_rat_2 SPYPGVQINEEFCQRLKDGTMRRAPELATPAIRHIMQSCWSGDPKARPAFSDLVEILGDL
FLT4_rat SPYPGVQINEEFCQRLKDGTMRRAPELATPAIRHIMQSCWSGDPKARPAFSDLVEILGDL
FLT4_mouse SPYPGVQINEEFCQRLKDGTMRRAPELATPAIRHIMQSCWSGDPKARPAFSDLVEILGDL
FLT4_human SPYPGVQINEEFCQRLKDGTMRRAPELATPAIRIMLCNWSGDPKARPAFSDLVEILGDL
VEGFR3_chicken SPYPGVQINEEFCQRLKDGTMRRAPEYATAEIYRIMLCSWHGDPKERPTFSDLVEILGNL
FLT1_mouse SPYPGVQMDEDFCSRLKEGMRMTPEYATPEIYQIMLCDCWHKDPKERPRFAELVEKLGLD
FLT1_rat SPYPGVQMDEDFCSRLKEGMRMTPEYATPEIYQIMLCDCWHKDPKERPRFAELVEKLGLD
FLT1_human SPYPGVQMDEDFCSRLREGMRRAPEYSTPEIYQIMLCDWHRDPERPRFAELVEKLGLD
FLT1_chicken SPYPGVQIDEDFCSKLKEGTRMRAPEQATEEIYQIMLCDWRSNPNERPWSELVKRLGDL
VEGFR2_chicken SPYPGVQIDEDFCRRLKEGTRMRSPSYSTPEVYTMLDCWGHGVPTERTFTTELVERIGDL
KDR_fly SPYPLGHIDEEFCRRLKEGTRMKAPESYSSEIYTMLDCWGHGPSQRTFTTELVERIGDL
*****:::***:::***:::***:::***:::***:::***:::***:::***:::***:::***:::***:::***
```

```

KDR_mouse      LQAN-AQQDGKDIYIVLPMSSETLSMEEDSGLSLPTSPVSCMEEEEVCDP-----KFHY
KDR_rat        LQAN-AQQDGKDIYIVLPMSSETLSMEEDSGLSLPTSPVSCMEEEEVCDP-----KFHY
KDR_human      LQAN-AQQDGKDIYIVLPISSETLSMEEDSGLSLPTSPVSCMEEEEVCDP-----KFHY
KDR_chicken    LQAN-VRQDGKDYVVLPLSPVSLNMEEDSGLSLPTSPASCKEEVEVCDP-----KFHY
FLT4_rat_2     LQGGGWQEEEEECMAL--HSSQSSEEDGFMQASTTALHITeADAESSPPSMHCHSLAARY
FLT4_rat       LQGGGWQEEEEECMAL--HSSQSSEEDGFMQASTTALHITeADAESSPPSMHCHSLAARY
FLT4_mouse     LQGGGWQEEEEERMAL--HSSQSSEEDGFMQASTTALHITeADAESSPPSMHCHSLAARY
FLT4_human     LQGRGLQEEEEVMAP--HSSQSSEEGFSQVSTMALHIAQADAEDSPSLQRHSLAARY
VEGFR3_chicken LQEN-VQEGKDYIPL--NDSHSEEDDGFSQVPSSAQ--QNSDEEDFDMIRCHSLAARY
FLT1_mouse     LQAN-VQQDGKDYIPL--NAILTRNSGTYSTPTFSEDLFKDGFADPHFH-----
FLT1_rat       LQAN-VQQDGKDYIPL--NAILTRNSGTYSTPTFSEDFDKDGTDPKHF-----
FLT1_human     LQAN-VQDGKDYIPI--NAILTNGSGTYSTPAFSEDFFKSISAPFN-----
FLT1_chicken   LQAS-VQEGKDYIPL--DTIFTAESGF---PPASDPLCNEKFPVPSPN-----
VEGFR2_chicken LQAN-VQQDGKDYIPL--NITLCPDGEPSKTCPEENLNNCVNRWSAVE-----
KDR_fly        LQAS-VQEGKHYIPI--NTALLTKADP-SNQSPTEETSTRPVS-LRDSG-----
**            :::::

```

-66-

```

KDR_mouse      DNTAGISHYLQNSKRKSRPVSVKTFEDIPLLEEPEVKVIPDDSQDSDGMVLASEELKTLED
KDR_rat        DNTAGISHYLQNSKRKSRPVSVKTFEDIPLLEEPEVKVIPDDSQDSDGMVLASEELKTLED
KDR_human      DNTAGISQYLQNSKRKSRPVSVKTFEDIPLLEEPEVKVIPDDNQDSDGMVLASEELKTLED
KDR_chicken    DNTAGISQYRQGSKRKSRPVSVKTFEDIPLVTT-VKVVQEEHQDSDGMVLASEELKTLEE
FLT4_rat_2     YNCVSFPGRLVRGRTKAPGSSRMKTFEELPMTPTTYKASVD-NQDSDGMVLASEEFEQIES
FLT4_rat       YNCVSFPGRLVRGRTKAPGSSRMKTFEELPMTPTTYKASVD-NQDSDGMVLASEEFEQIES
FLT4_mouse     YNCVSFPGRLARGTKTPGSSRMKTFEELPMTPTTYKASMD-NQDSDGMVLASEEFEQIES
FLT4_human     YNWVSFPGCLARGAETRGSSRMKTFEELPMTPTTYKGSVD-NQDSDGMVLASEEFEQIES
VEGFR3_chicken YNCVSFPGCLTGGNQIRCSRIKTFEELPMTHTMYKAHPD-NQDSDGMVLASEEFEQIES
FLT1_mouse     -SGSSDDVRYVNAFKFMSLERIKTFEELSPNS---TSMFEDYQLDSTLLGSPLLKRFTW
FLT1_rat       -SGSSDDVRYVNAFKFMSLERIKTFEELSPNA---TSMFEDYHLDTSSLLTSPLLKRFTW
FLT1_human     -SGSSDDVRYVNAFKFMSLERIKTFEELLPNA---TSMFDDYQDSSSTLLASPLMKRFTW
FLT1_chicken   -CRSTERARYINTFKIKPPQRIKTFEELPIKE---KLVFNQYQADSDGMVLASEELKRFTW
VEGFR2_chicken -TSGSGSKRR-----PLSVKTFDEVSVGK--EKVMHE--ESDSGMVLTSDEIKSPKR
KDR_fly        -TAWNIR-----PESVKTFFDEVILENG-TNKIHEGGQSDSGIGLSSDDLKTLKR

```

```

:***:: . . *.. * * ::

```

```

KDR_mouse      RNK-----LSPSFGGMMPSKSRRESVASEGSNQTSQGYQSGYHSDDTDTTVYSSDEAGLLKM
KDR_rat        RNK-----LSPSFGGMMPSKSRRESVASEGSNQTSQGYQSGYHSDDTDTTVYSSDEAGLLKL
KDR_human      RTK-----LSPSFGGMMPSKSRRESVASEGSNQTSQGYQSGYHSDDTDTTVYSSDEAEALLKL
KDR_chicken    RDKQ----VKIPFSTLAPSKSNESVMSEASNQTSQGYQSGYHSDMDTMVCSSEDTELLCA
FLT4_rat_2     RHR-----QEGSFSRKDPGQHMD-ISRGHDPDLQGRRRRPTQGAQGGKVFNNEYEVSQP
FLT4_rat       RHR-----QEGSFSRKDPGQHMD-ISRGHDPDLQGRRRRPTQGAQGGKVFNNEYEVSQP
FLT4_mouse     RHR-----PEGSFCKGPGQHMD-IPRGHPDPQGRRRRPTQGAQGGKVFNNEYEVSQP
FLT4_human     RHR-----QESGFR-----
VEGFR3_chicken RHR-----KEGGFSSKGNRTAE-LSAEQSDLRGCRPSYGSQVGGQTFFYNSEYGESEH
FLT1_mouse     TETK--PKASMKIDLRIASKSKEAGLSDLPRP-SFCFSSCGHIR-PVQ---DDES--ELG
FLT1_rat       TETK--PKASMKIDLRIASKSKEAGLSDLPGP-SFCFSSCGHIR-PVRQEDDDP--ELG
FLT1_human     TDSK--PKASLKIDLRVTSKSKESGLSDVSRP-SFCHSSCGHVSEGKRRTYDHA--ELE
FLT1_chicken   TGSK--QKWTL-FGMKGVSRSKESGLSGITKPRSFCFSFCDQLSESKRRYTYGNT--VLE
VEGFR2_chicken LEIRSWPYGIMALARRAVNKSILSDHEREAVKYQPAVQVEEDTMDFFLEDVLLPMD
KDR_fly        LESLARPRSFMSRAMK--RKSKEVSL--LEGEMDKYPLVP-----SLSLEDS---SLD

```

KDR_mouse	VDAAVHA-----DSGTTTLRSPFV	SEQ ID NO:19
KDR_rat	VDVAGHV-----DSGTTTLRSPFV	SEQ ID NO:20
KDR_human	IEIGVQTGSTAQILQPDSTLSSPPV	SEQ ID NO:21
KDR_chicken	QEASPTLPRVHG-LVHDSAPPLVSPPL	SEQ ID NO:22
FLT4_rat_2	CTEGDCCP-----SAGSTFFADSNY	SEQ ID NO:23
FLT4_rat	CTEGDCCP-----SAGSTFFADSNY	SEQ ID NO:24
FLT4_mouse	CTEGDCCP-----SAGSTFFADSSY	SEQ ID NO:25
FLT4_human	-----	SEQ ID NO:26
VEGFR3_chicken	SEDGSCPTPPGEGASPTALHTSFSEQY	SEQ ID NO:27
FLT1_mouse	-KESCCSPPPD----YNSVVLYSSPPA	SEQ ID NO:28
FLT1_rat	-KESCCSPPPD----YNSVVLYSSPPA	SEQ ID NO:29
FLT1_human	RKIACCSPPPD----YNSVVLYSTPPI	SEQ ID NO:30
FLT1_chicken	KMKACHSPPPD----YSSVVHYQPSI	SEQ ID NO:31
VEGFR2_chicken	PSLECHSPPPD----YNSVVQCSAPPV	SEQ ID NO:32
KDR_fly	SEMECHSPPPD----YNYVVRYSTPPI	SEQ ID NO:33

[0187] A summary of the sequence alignment of FLT-4 sequences from multiple organisms is shown in TABLE 20 below.

TABLE 20
Summary of Sequence Alignment of FLT-4 Sequences from Multiple Organisms

<u>Mutation</u>	<u>Comment</u>
N149D	Conserved in FLT4 and KDR
M963V	Conserved in FLT4
D970E	Conserved in FLT4 and KDR
P1114S	Conserved in FLT4, FLT1 and KDR

[0188] *Analysis of the Potential Effect of FLT-4 mutations on FLT-4 Protein Secondary Structure. The Effect of FLT-4 Mutation on Amino Acid Property.* The change of amino acid property observed by FLT-4 mutation is summarized in TABLE 21 (Valdar WS, *Proteins* 48(2): 227-41 (2002)).

TABLE 21
Influence of FLT-4 Mutations on FLT-4 Amino Acid Property

<u>Mutation</u>	<u>Property change</u>
N149D	None predicted.
M963V	None predicted.
D970E	None predicted.
P1114S	None predicted.

[0189] *nnPredict Method Analysis of the Wild-type FLT-4 Secondary Structure.* Secondary structure predictions of wild-type FLT-4 (TABLE 22) and mutant FLT-4 polypeptides (TABLE 24; TABLE 26; TABLE 28; and TABLE 30) were performed by nnPredict. The basis of the prediction is a two-layer, feed-forward neural network. The network weights were determined by a separate program -- a modification of the Parallel Distributed Programming suite of McClelland and Rumelhart (MIT Press, Cambridge MA.1, Vol. 3, pp 318-362 (1988)). Complete details of the determination of the network weights is found in Kneller *et. al.* (*J. Mol. Biol.*, (214): 171-182 (1990)). The output is a secondary structure prediction for each position in the sequence.

[0190] All TABLES (*e.g.*, TABLE 23; TABLE 25; TABLE 27; TABLE 29; and TABLE 31) in this section that summarize the FLT-4 protein secondary structure as predicted by nnPredict use "H", "E" and a dash "-" as identifiers, which are defined as follows. A helix element is designated by the letter "H". A strand element is designated by the letter "E". No prediction is designated by a dash ("-"). Grey shading represents polypeptide regions where mutation was identified.

-68-

[0191] The amino acid sequence of wild-type FLT-4 polypeptide (SEQ ID NO:26) is shown below in TABLE 22. The position of the mutated amino acid residues is highlighted in bold, underlined text.

TABLE 22

MQGAALCLRLWLCGLLDGLVSGYSMTPTLNITEESHVIDTGDSLISCRGQHPLEWA
 WPGAQEAPATGDKDSED TG VVRDCEGTDARPYCKVLLLHEVHANDTGSYVCYKYIKARI
 EGTAAASSYVFVRDFEQPFINKPDTLLV**NR**KDAMWVPCLVSI PGLNVTLRSQSSVLWPDG
 QEVVWDDRRGMLVSTPLLHDALYLQCETTWGDQDFLSNPFLVHITGNELYDIQLLPRKSL
 ELLVGEKLVLNCTVWAEFNSGVTFDWDYPGKQAERGKWVPERRSQQTHTELSSILTIHNV
 SQHDLGSYVCKANNGIQRFRESTEVI VHENPFISVEWLKGPILEATAGDELVKLPVKLAA
 YPPPEFQWKDGKALSGRHSFHALVLKEVTEASTGT YTLALWNSAAGLRRNISLELVNV
 PPQIHEKEASSPSIYSRHSRQALTCTAYGVPLPLSIQWHWRPWT PCKMFAQRSLRRRQQQ
 DLMPQCRDWRAVTTQDAVNPIESLDTWTEFVEGKNKTVSKLVIQANVVSAMYKCVVSNKV
 GQDERLIYFYVTTIPDGFTIESKPSEELLEGPVLLSCQADSYKYEHLRWYRLNLSTLHD
 AHGNPLLLDCKNVHLFATPLAASLEEVAPGARHATLSLSIPRVAPEHEGHYVCEVQDRRS
 HDKHCHKYLSVQALEAPRLTQNLTDLLVNVSDSLEMQLVAGAHAPSIVWYKDERLLEE
 KSGVDLADSNQKLSIQRVREEDAGRYLCSVCNAKGCVNSSASVAVEGSEDKGSMEIVILV
 GTGVIAVFFWVLLLLIFCNMRPAHADIKTG YLSIIMDPGEVPLEEQCEYLSYDASQWEF
 PRERLHLGRVLGYGAFGKVVEASAFGIHKGSSCDTVAVKMLKEGATASEHRALMSELKIL
 IHIGNHLNVVNLLGACTKPQGPLMVIVEFCKYGNLSNFLRAKRDAFSPCAEKSPQGRGF
RAMVELAR**L**RRRRPGSSDRVLFARFSKTEGGARRASPDQEAEDLWLSPLTMEDLVCYSFQ
 VARGMEFLASRKCIHRDLAARNILLSESDVVKICDFGLARDIYKDPDYVRKGSARLPLKW
 MAPESIFDKVYTTQSDVWSFGVLLWEIFSLGAS**P**YPGVQINEEFCQRLRDGTRMRAPELA
 TPAIRRIMLNCWSGDPKARPAFSELVEILGDLLQGRGLQEEEEVCMAPRSSQSSEEGSFS
 QVSTMALHIAQADAEDSPPSLQRHSLAARYYNWVSFPGCLARGAETRGS SRMKT FEEFPM
 TPTTYKGSVDNQTDSGMVLASEEFEQIESRHRQESGFR

SEQ ID NO:26

-69-

[0192] A schematic representation of the secondary structure of wild-type FLT-4 polypeptide (SEQ ID NO:26) predicted using nnPredict analysis is shown below in TABLE 23. The position of the mutated amino acid residues are identified with grey shaded text.

TABLE 23

```

-----HHHHHHHHHHH-----E-----EEE-----EEE-----
-----EEE-----HHHHHHHHH-----EEEEH-----EEEE
E-----EEEE-----EH-----EEF-----
--EE-----EE--HH-HHHHE-----EEEE--HHH--H--H
HHHHH-HHEEH-EEHE--E-----E--EEEEEE--
-----EEH-----EEE-----EEE-----EEH--HHHHHHH--
-----EHHHHHHH-----EEHHHHHHH-H-H-----EEE--
-----EEE-----EEE-----E-E-----HHH-----
-----HHHHH-----HEHEH-H--HHHEE-----
--HHEEEEE-----EE-----HHH-----EEE-----HHHHHHHHH-HH-H--
-----HH-----HH-HHHHHHHHH-HH-HHH-----EEE-----
-----H--HHH-H--HHHH--HHHEE-----HHHHHHH-----EEEH--HHHH--
-----HH-----EE-----HEEE-----E-----EEE-----EEEEEE
--EEEEHHHHHHHHHHH-----EEEE-----HHHHHH--
--HHHH--EE-----EHEHHHHHE-----HHHHHHHHH--HHHHHHHHHHHHE
EHH--HHHEH-----EEEEHH-----HHHHH-----
HHHHHHH--HEHEE-----HHHHH-----HHHHHHHH
--HHHHHHH--HH-----HHHHHH-----EEEH-----H-----H--
-----EE-----EEH-HEEEHHH-----HHHHHH-----
--HHHHHHHEE-----HHHHHHHHHHH-----HHHHH-----
EHHHHHHHHH-----HH-HHHHHHEE-----
-----EEEH--HHHHHH-----

```

-70-

[0193] The amino acid sequence of FLT-4 mutant polypeptide N149D (SEQ ID NO:34) is shown below in TABLE 24. The position of the mutated amino acid residue is highlighted in bold underlined text.

TABLE 24

MQRGAALCLRLWLCLGLLDGLVSGYSMTPTLNITEESHVIDTGDSLISISCRGQHPLEWA
 WPGAQEAPATGDKDSEDTGVVRDCEGTDARPYCKVLLLHEVHANDTGSYVCYKYIKARI
 EGTAAASSYVFVRDFEQPFINKPD~~LLVDR~~KDAMWVPCLVSI PGLNVTLRSQSSVLWPDG
 QEVVWDDRRGMLVSTPLLHDALYLQCE~~T~~TWGDQDFLSNPFLVHITGNELYDIQLLPRKSL
 ELLVGEKLVLNCTVWAEFNSGVTFDWDYPGKQAERGKWVPERRSQQTHTELSSILTIHNV
 SQHDLGSYVCKANNGIQRFRESTEVI~~V~~HENPFISVEWLKGP~~I~~EATAGDELVKLPVKLAA
 YPPPEFQWKDGKALSGRHS~~P~~HALVLKEVTEASTGTYTLALWNSAAGLRRNISLELVNV
 PPQIHEKEASSPSIYSRHSRQALTCTAYGVPLPLSIQWHWRPWPCKMFAQRS~~L~~RRRQQQ
 DLMPQCRDWRAVTTQDAVNPIESLDTWTEFVEGKNKTVSKLVIQ~~N~~ANVSAMYKCVVSNKV
 GQDERLIYFYVTTIPDGFTIESKPS~~E~~ELLEGGQPVLLSCQADSYKYEHLRWYRLNLSTLHD
 AHGNPLLLDCKNVHLFATPLAASLEE~~V~~APGARHATLSLSIPRVAPEHEGHYVCEVQDRRS
 HDKHCHKYLSVQALEAPRLTQNLTDLLVNVS~~D~~SLEMQCLVAGAHAPSIVWYKDERLLEE
 KSGVDLADSNQKLSIQRVREEDAGRYLCSVCNAKGCVN~~S~~SASVAVEGSEDKGSMEIVILV
 GTGVIAVFFWVLLLLIFCNMRRPAHADIKTYLSIIMDPGEVPLEEQCEYLSYDASQW~~E~~F
 PRERLHLGRVLGYGAFGKVVEASAFGIHKGSSCDTVAVKMLKEGATASEHRALMSELKIL
 IHIGNHLNVVNLLGACTKPQGPLMVIVEFCKYGNLSN~~F~~LRAKRD~~A~~FSPCAEKSP~~EQRGRF
 RAMVELARLDRRRPGSSDRVLFARFSKTEGGARRASPDQEAEDLWLSPLTMEDLVCYSFQ
 VARGMEFLASRKCIHRDLAARNILLSESDVVKICDFGLARDIYKDPDYVRKGSARLPLKW
 MAPESIFDKVYTTQSDVWSEGVLLWEI~~F~~SLGASPYPGVQINEEFCQRLRDGTRMRAPELA
 TPAIRRIMLNCWSGDPKARPAFSELVEILGDLLQGRGLQ~~E~~EEVCMAPRSSQSSEEGSFS
 QVSTMALHIAQADAEDSPPSLQRHSLAARYYNWVSFPGCLARGAETRGSSRMKT~~F~~EEFPM
 TPTTYKGSVDNQTDSGMVLASEE~~F~~EQIESRHRQESGFR~~

SEQ ID NO:34

-71-

[0194] A schematic representation of the secondary structure of FLT-4 mutant polypeptide N149D (SEQ ID NO:34) predicted using nnPredict analysis is shown below in TABLE 25.

The position of the mutated amino acid residue is identified by grey shaded text.

TABLE 25

```

-----HHHHHHHHHHH-----E-----EEE-----EEE-----
-----EEE-----HHHHHHHHH-----EEEEEH-----EEEE
E-----EEEE-----EH-----EEE-----
--EE-----EE--HH-HHHHE-----EEEE-----HHHH--H--H
HHHHH-HHEEEH-EEHE-----E-----E-----EEEEEE--
-----EEH-----EEE-----EEEE-----EEE-----EEH--HHHHHHHH--
-----EHHHHHHH-----EEHHHHHHHH-H-H-----EEE--
-----EEE-----EEE-----E-E-----HHH-----
-----HHHHH-----HEHEH-H--HHHEEE-----
--HHEEEEEEE-----EE-----HHH-----EEE-----HHHHHHHHHH-HH-H--
-----HH-----HH-HHHHHHHHH-HH-HHH-----EEE-----
-----H--HHH-H--HHHH--HHHEE-----HHHHHHH-----EEEH--HHHH-
-----HH-----EE-----HEEE-----E-----EEE-----EEEEEEE
--EEEEHHHHHHHHHHH-----EEEE-----HHHHHH-----
--HHHH--EE-----EHEHHHHHHE-----HHHHHHHHH--HHHHHHHHHHHHE
EHH---HHHEH-----EEEEHH-----HHHHH-----
HHHHHHHHH-----HEHEE-----HHHHH-----HHHHHHHH
--HHHHHHH--HH-----HHHHHH-----EEEH-----H-----H--
-----EE-----EEH-HEEEHHH-----HHHHHH-----
-HHHHHHEE-----HHHHHHHHHHH-----HHHHH-----
EHHHHHHHHH-----HH-HHHHHEE-----
-----EEEH--HHHHHH-----

```

-72-

[0195] The amino acid sequence of FLT-4 mutant polypeptide M963V (SEQ ID NO:35) is shown below in TABLE 26. The position of the mutated amino acid residue is highlighted in bold underlined text.

TABLE 26

MQGAALCLRLWLCGLLDGLVSGYSMTPTLNITEESHVIDTGDSLISISCRGQHPLEWA
 WPGAQEAPATGDKDSED TGVRDCEGT DARPYCKVLLLHEVHANDTGSYVCYKYIKARI
 EGT TAASSYVFVRDFEQPFINKPDTLLVNRKDAMWVPCLVSI PGLNVTLRSQSSVLWPDG
 QEVVWDDRRGMLVSTPLLHDALYLQCETTWGDQDFLSNPFVLVHITGNELYDIQLLPRKSL
 ELLVGEKLVLNCTVWAEFNSGVTFDWDYPGKQAERGKWPERRSQQTHTELSSILTIHNV
 SQHDLGSYVCKANNGIQRFRESTEVI VHENPFISVEWLKGP ILEATAGDELVKLPVKLAA
 YPPPEFQWYKDGKALSGRHS PHALVLKEVTEASTGT YTLALWNSAAGLRNISLELVNV
 PPQIHEKEASSPSIYSRHSRQALTCTAYGVPLPLSIQWHWRPWPCKMFAQRS LRRRQQQ
 DLMPQCRDWRAVTTQDAVNPIESLDTWTEFVEGKNKTVSKLVIQANVVSAMYKCVVSNKV
 GQDERLIYFYVTTIPDGFTIESKPSEELLEGPVLLSCQADSYKYEHLRWYRLNLSTLHD
 AHGNPLLLDCKNVHLFATPLAASLEEVAPGARHATLSLSIPRVAPEHEGHYVCEVQDRRS
 HDKHCHKKYL SVQALEAPRLTQNLTDLLVNVSDSLEMQCLVAGAHAPSIVWYKDERLLEE
 KSGVDLADSNQKLSIQRVREEDAGRYLCSVCNAKGCVNSSASVAVEGSEDKGSMEIVILV
 GTGVI AVFFFWVLLLLIFCNMRRPAHADIKTG YLSIIMDPGEVPLEEQCEYLSYDASQWEF
 PRERLHLGRVLGYGAFGKVVEASAFGIHKGSSCDTVAVKMLKEGATASEHRALMSELKIL
 IHIGNHLNVVNLLGACTKPGPLMVIVEFCKYGNLSNFLRAKRDAFSPCAEKSPEQRGRF
 RAVVVELARLD RRRPGSSDRVLFARFSKTEGGARRASPDQEAEDLWLSPLTMEDLVCYSFQ
 VARGMEFLASRKCIHRDLAARNILLS ESDVVKICDFGLARDIYKDPDYVRKGSARLPLKW
 MAPESIFDKVYTTQSDVWSFGVLLWEIFSLGASPYPGVQINEEFCQRLRDGTRMRAPELA
 TPAIRRIMLNCWSGDPKARPAFSELVEILGDLLQGRGLQEEEEVCMAPRSSQSSEEGSFS
 QVSTMALHIAQADAEDSPPSLQRHSLAARYYNWVSFPGCLARGAETRGS SRMKT FEEFPM
 TPTTYKGSVDNQTDSGMVLASEEFEQIESRHRQESGFR

SEQ ID NO:35

-73-

[0196] A schematic representation of the secondary structure of FLT-4 mutant polypeptide M963V (SEQ ID NO:35) predicted using nnPredict analysis is shown below in TABLE 27. The position of the mutated amino acid residue is identified by grey shaded text.

TABLE 27

```

-----HHHHHHHHHHHH-----E-----EEE-----EEE-----
-----EEE-----HHHHHHHHHH-----EEEEEH--EEEE
E-----EEEE-----EH-----EEE-----
--EE-----EE--HH-HHHHE-----EEE-----HHHH--H--H
HHHHH-HHEEEH-EEHE-----E-----E-EEEEEE--
-----EEH-----EEE-----EEEE-----EEE-----HHHHHHHH--
-----EHHHHHHH-----EEHHHHHHHH-H-H-----EEE--
-----EEE-----EEE-----E-E-----HHH-----
-----HHHHH-----HEHEH-H--HHHEEE-----
--HHEEEEEE-----EE-----HHH-----EEE-----HHHHHHHHHH-HH-H--
-----HH-----HH-HHHHHHHHH-HH-HHH-----EEE-----
-----H--HHH-H--HHHK--HHHEE-----HHHHHHHH-----EEEH--HHHH-
-----HH-----EE-----HEEE-----E-----EEE-----EEEEEE
--EEEEHHHHHHHHHHHH-----EEE-----HHHHHH-----
--HHHH--EE-----EHHHHHHHHE-----HHHHHHHHHH--HHHHHHHHHHHHHE
EHH--HHHEH-----EEEEHH-----HHHHH-----
EHHHHHHH-----HEHEE-----HHHHH-----HHHHHHHH
--HHHHHHHH--HH-----HHHHHH-----EEEH-----H-----H--
-----EE-----EEH-HEEEHH-----HHHHHH-----
--HHHHHEEE-----HHHHHHHHHHHH-----HHHHH-----
EHHHHHHHHHH-----HH-HHHHHHEE-----
-----EEEH--HHHHHH-----

```

-74-

[0197] The amino acid sequence of FLT-4 mutant polypeptide D970E (SEQ ID NO:36) is shown below in TABLE 28. The position of the mutated amino acid residue is highlighted in bold underlined text.

TABLE 28

MQGAALCLRLWLCLGLLDGLVSGYSMTPTLNITEESHVIDTGDSLISICRGQHPLEWA
 WPGAQEAPATGDKDSED TG VVRDCEGT DARPYCKVLLLHEVHANDTGSYVCYKYIKARI
 EGT TAASSYVFVRDFEQPFINKPD TLLVNRKDAMWVPCLV SIPGLNVTLRSQSSVLWPDG
 QEVVWDDRRGMLVSTPLLH DALYLQ CETTWGDQDFLSNPFLVHITGNELYDIQLLPRKSL
 ELLVGEKLV LNCTVWAEFNSGVTFDWDYPGKQAERGKWPERRSQOHTHELSSILTIHNV
 SQHDLGSYVCKANNGIQRFRESTEVI VHENPFISVEWLKGP ILEATAGDELVKLPVKLAA
 YPPPEFQWYKDGKALSGRHS PHALVLKEVTEASTGTYTLALWNSAAGLRN ISLELVNV
 PPQIHEKEASSPSIYSRHSRQALTCTAYGVPLPLSIQWHWRP WTPCKMFAQRSLRRRQQQ
 DLMPQCRDWRAVTTQDAVNPIESLDTWTEFVEGKNKTVSKLVIQANVNSAMYKCVVSNKV
 GQDERLIYFYVT TIPDGFTIESKPSEELLEGQPVLLSCQADSYKYEHLRWYRLNLSTLHD
 AHGNPLLLDCKNVHLFATPLAASLEEVAPGARHATLSLSI PRVAPEHEGHYVCEVQDRRS
 HDKHCHKKYL SVQALEAPRLTQNLTDLLVNVSDSLEMQCLVAGAHAPSIVWYKDERLLEE
 KSGVDLADSNQKLSIQRVREEDAGRYLCSVCNAKGCVNSSASVAVEGSEDKGSM EIVILV
 GTGVIAVFFWVLLLLIFCNMRPAHADIKTG YLSIIMDPGEVPLEEQCEYLSYDASQWEEF
 PRERLHLGRVLGYGAFGKVVEASAFGIHKGSSCDTVAVKMLKEGATASEHRALMSELKIL
 IHIGNHLNVVNLLGACTKPGQPLMVIVEFCKYGNLSNFLRAKRD AFSPCAEKSPEQRGRF
 RAMVELARLE~~ERRR~~PGSSDRVLFARFSKTEGGARRASPDQEAEDLWLSPLTMEDLVCYSFQ
 VARGMEFLASRKCIHRDLAARNILLS ESDVVKICDFGLARDIYKDPDYVRKGSARLPLKW
 MAPESIFDKVYTTQSDVWSFGVLLWEIFSLGASPYPGVQINEEFCQRLRDGTRMRAPELA
 TPAIRRIMLNCWSGDPKARPAFSELVEILGDLLQGRGLQEEEEVCMAPRSSQSSEEGSFS
 QVSTMALHIAQADAEDSPPSIQRHSLAARYYNWVSFPGCLARGAETRGSSRMKT FEEFPM
 TPTTYKGSVDNQTDSGMVLASEEFEQIESRHRQESGFR

SEQ ID NO:36

[0198] A schematic representation of the secondary structure of FLT-4 mutant polypeptide D970E (SEQ ID NO:36) predicted using nnPredict analysis is shown below in TABLE 29.

The position of the mutated amino acid residue is identified by grey shaded text.

TABLE 29

[illegible]

-76-

[0199] The amino acid sequence of FLT-4 mutant polypeptide P1114S (SEQ ID NO:37) is shown below in TABLE 30. The position of the mutated amino acid residue is highlighted in bold underlined text.

TABLE 30

MQRGAALCLRLWLCLGLLDGLVSGYSMTPTLNITEESHVIDTGDSLISISCRGQHPLEWA
 WPGAQEAPATGDKDSEDTGVVRDCEGTDARPYCKVLLLHEVHANDTGSYVCYYKYIKARI
 EGTTAASSYVFVRDFEQPFINKPDTLVNRKDAMWVPCLVSIPLGNVTLRSQSSVLWPDG
 QEVVWDDRRGMLVSTPLLHDALYLQCETTWDQDFLSNPFLVHITGNELYDIQLLPKSL
 ELLVGEKLVLNCTVWAEFNSGVTFDWDYPGKQAERGKWPERRSQQTHTELSSILTIHNV
 SQHDLGSYVCKANNGIQRFRESTEVI VHENPFISVEWLKGPILEATAGDELVKLPVKLAA
 YPPPEFQWYKDGKALSGRHSFHALVLKEVTEASTGTYTLALWNSAAGLRRNISLELVNV
 PPQIHEKEASSPSIYSRHSRQALTCTAYGVPLPLSIQWHWRPWT PCKMFAQRSLRRRQQQ
 DLMPQCRDWRAVTTQDAVNPIESLDTWTEFVEGKNKTVSKLVIQNANVSAMYKCVVSNKV
 GQDERLIYFYVTTIPDGFTIESKPSSEELLEGPVLLSCQADSYKYEHLRWYRLNLSTLHD
 AHGNPLLLDCKNVHLFATPLAASLEEVA PGARHATLSLSIPRVAPEHEGHYVCEVQDRRS
 HDKCHCHKYLSVQALEAPRLTQNLTDLLVNVSDSLEMQCLVAGAHAPSIVWYKDERLLEE
 KSGVDLADSNQKLSIQRVREEDAGRYLCSVCNAKGCVNSSASVAVEGSEDKGSMEIVILV
 GTGVIAVFFFWVLLLLIFCNMRRPAHADIKTGYSIIIMDPGEVPLEEQCEYLSYDASQWEF
 PRERLHLGRVLGYGAFGKVVEASAFGIHKGSSCDTVAVKMLKEGATASEHRALMSELKIL
 IHIGNHLNVVNLLGACTKPQGPLMVIVEFCKYGNLSNFLRAKRDAFSPCAEKSPEQRGRF
 RAMVELARLD RRRPGSSDRVLFARFSKTEGGARRASPDQEAEDLWLSPLTMEDLVCYSFQ
 VARGMEFLASRKCIHRDLAARNILLSSESDVVKICDFGLARDIYKDPDYVRKGSARLPLKW
 MAPESIFDKVYTTQSDVWSFGVLLWEIFSLGASSYPGVQINEEFCQRLRDGTRMRAPELA
 TPAIRRIMLNCWSGDPKARPAFSELVEILGDLLQGRGLQEEEEVCMAPRSSQSSEEGSFS
 QVSTMALHIAQADAEDSPPSLQRHSLAARYYNWVSFPGCLARGAETRGS SRMKT FEEFPM
 TPTTYKGSVDNQTDSGMVLASEEFEQIESRHRQESGFR

SEQ ID NO:37

-77-

[0200] A schematic representation of the secondary structure of FLT-4 mutant polypeptide P1114S (SEQ ID NO:37) predicted using nnPredict analysis is shown below in TABLE 31. The position of the mutated amino acid residue is identified by grey shaded text.

TABLE 31

```

-----HHHHHHHHHHHH-----E-----EEE-----EEE-----
-----EEE-----HHHHHHHHHH-----EEEEEH-----EEEE
E-----EEEE-----EH-----EEE-----
--EE-----EE--HH-HHHHE-----EEEE-----HHHH--H--H
HHHHH-HHEEEH-EEHE-----E-----E-----EEEEEE--
-----EEH-----EEE-----EEE-----EEH-----HHHHHHHH--
-----EHHHHHHHH-----EEHHHHHHHH-H-H-----EEE--
-----EEE-----EEE-----E-E-----HHH-----
-----HHHHH-----HEHEH-H--HHHEEE-----
--HHEEEEEEE-----EE-----HHH-----EEE-----HHHHHHHHHH-HH-H--
-----HH-----HH-HHHHHHHHH-HH-HHH-----EEE-----
-----H--HHH-H--HHHH--HHHEE-----HHHHHHHH-----EEEH--HHHH--
--HH-----EE-----HEEE-----E-----EEE-----EEEEEEE
--EEEEHHHHHHHHHHHH-----EEE-----HHHHHH-----
--HHHH--EE-----EHEHHHHHHE-----HHHHHHHHHH--HHHHHHHHHHHHE
EHH--HHHEH-----EEEEHH-----HHHHH-----
HHHHHHHH-----HEHEE-----HHHHH-----HHHHHHHH
--HHHHHHHH--HH-----HHHHHH-----EEEH-----H-----H--
-----EE-----EEH-HEEEHHHH-----E-HHHHHH-----
-HHHHHHEE-----HHHHHHHHHH-----HHHHH-----
EHHHHHHHHHH-----HH-HHHHHEE-----
-----EEEH--HHHHHH-----

```

[0201] The influence of FLT-4 mutations on FLT-4 protein secondary structure is summarized below in TABLE 32.

TABLE 32

Influence of FLT-4 Mutations on FLT-4 Protein Secondary Structure

<u>Mutation</u>	<u>Predicted Protein Secondary Structure Change</u>
N149D	No change
M963V	Affects 1 amino acid
D970E	Extend upstream helix by 2 amino acids
P1114S	Affects structure

-78-

[0202] *General Overview Analysis.* A summary of the results of computational analysis of the effect of the FLT-4 mutations and SNPs identified in the present invention on select features of wild-type FLT-4 is provided below in TABLE 33.

TABLE 33
Evaluation of FLT-4 Mutations by Sequence Features

<u>Mutation</u>	<u>Protein domain</u>	<u>Phosphorylation</u>	<u>Other modification</u>	<u>AA conservation</u>	<u>AA property change</u>	<u>Secondary Structure</u>
N149D				++		
M963V	++			+		+
D970E	++			++		++
P1114S	++	+	+	+++		+

+: the effect of mutation on protein function is low
 ++: the effect of mutation on protein function is medium
 +++: the effect of mutation on protein function is high

[0203] Four FLT-4 mutations were identified in colorectal cancer. The N149D had been identified previous as a SNP. M963V, D970E and P1114S were found in the protein tyrosine kinase domain. P1114 was highly conserved in FLT-4 as well as homologues, FLT1 and KDR. D970E was predicted to affect the protein secondary structure by extending an upstream helix in FLT-4 polypeptide. M963V and P1114S were also predicted to alter the polypeptide's secondary structure.

EXAMPLE II

ANALYSIS OF FLT-4 MUTATION FOR THERANOSTIC CANCER TREATMENT IN A SUBJECT

[0204] In this invention, an agent that modulates FLT-4 biological activity (*i.e.*, FLT-4 modulating agent, *e.g.*, FLT-4 antagonist) is administered to a patient with cancer, *e.g.*, colon cancer, when the patient has a single nucleotide polymorphism (SNP) pattern indicative of an FLT-4 mutation that correlates with the disease. In one embodiment, the SNP is selected from the group consisting of the FLT-4 mutation summarized in TABLE 1 and TABLE 2.

[0205] .

EQUIVALENTS

[0206] The present invention is not to be limited in terms of the particular embodiments described in this application, which are intended as single illustrations of individual aspects of the invention. Many modifications and variations of this invention can be made without departing from its spirit and scope, as will be apparent to those skilled in the art. Functionally equivalent methods and apparatuses within the scope of the invention, in addition to those enumerated herein, will be apparent to those skilled in the art from the foregoing descriptions. Such modifications and variations are intended to fall within the scope of the appended claims. The present invention is to be limited only by the terms of the appended claims, along with the full scope of equivalents to which such claims are entitled.

-80-

CLAIMS

We claim:

1. The use of an FLT-4 modulating agent in the manufacture of a medicament for the treatment of cancer in a selected patient population, wherein the patient population is selected on the basis of the genotype of the patients at a FLT-4 genetic locus indicative of a propensity for having cancer.
2. The use of an anti-cancer agent in the manufacture of a medicament for the treatment of cancer in a selected patient population, wherein the patient population is selected on the basis of the genotype of the patients at a FLT-4 genetic locus indicative of a propensity for having cancer.
3. An isolated polynucleotide having a sequence encoding an FLT-4 mutation listed in TABLE 1.
4. A vector comprising a polynucleotide of claim 3.
5. An organism containing a polynucleotide of claim 3.
6. The polynucleotide of claim 3, further comprising a polynucleotide sequence encoding an FLT-4 polypeptide having a sequence selected from the group consisting of: SEQ ID NO:34; SEQ ID NO:35; SEQ ID NO:36; and SEQ ID NO:37.
7. An isolated polypeptide having a sequence selected from the group consisting of: SEQ ID NO:34; SEQ ID NO:35; SEQ ID NO:36; and SEQ ID NO:37.
8. A method for treating cancer in a subject, comprising the steps of:
 - (a) obtaining the genotype or haplotype of a subject at a FLT-4 gene locus, wherein the genotype and/or haplotype is indicative of a propensity for having cancer; and
 - (b) administering an anti-cancer therapy to the subject.

-81-

9. The method of claim 8, wherein the anti-cancer therapy is selected from the group consisting of Glivec®, FEMARA®, Sandostatin® LAR®, ZOMETA®, vatalanib, everolimus, gimatecan, patupilone, midostaurin, pasireotide, LBH589, AEE788 and AMN107.
10. The method of claim 8, wherein the cancer is selected from the group consisting of: glioblastoma; breast cancer; melanoma, ovarian cancer, cholangioma; non-small-cell lung cancer (NSCLC); prostate cancer; and colon cancer.
11. The method of claim 8, wherein the genotype is heterozygous, with at least one of the alleles containing an FLT-4 polymorphism and/or mutation of TABLE 1.
12. The method of claim 8, wherein the genotype is homozygous, with at least one of the alleles containing an FLT-4 mutation or polymorphism of TABLE 1.
13. The method of claim 8, wherein the anti-cancer therapy is the administration of a therapeutically effective amount of an FLT-4 modulating agent.
14. A method for diagnosing a propensity for having cancer in a subject, comprising the steps of:
 - (a) obtaining the genotype or haplotype of a subject at a FLT-4 gene locus, wherein the genotype and/or haplotype is indicative of a propensity of the cancer to respond to the drug; and
 - (b) identifying the subject as having a propensity for having cancer.
15. The use of a polynucleotide having a sequence encoding an FLT-4 mutation listed in TABLE 1 as a drug target.