

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
13 April 2006 (13.04.2006)

PCT

(10) International Publication Number
WO 2006/038904 A2

(51) International Patent Classification:

G01N 33/53 (2006.01) **G01N 33/68** (2006.01)
G01N 33/574 (2006.01)

(21) International Application Number:

PCT/US2004/025269

(22) International Filing Date: 6 August 2004 (06.08.2004)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant (for all designated States except US): **AP-PLERA CORPORATION** [US/US]; Wayne W. Montgomery, Vice President, Intellectual Property, Celera Genomics, 45 West Gude Drive, CI-1#316, Rockville, MD 20850 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **DOMON, Bruno** [CH/US]; 440 Elmcroft Boulevard, Apt. 3410, Rockville, MD 20850 (US). **MCCAFFERY, Ian** [GB/US]; 104 Ladyshire Lane, C203, Rockville, MD 20850 (US). **NARAYAN, vaihhav** [IN/US]; 11617 Pinoy Lodge Road, North Potomac, MD 20878 (US). **PATTERSON, Scott** [AU/US]; 1003 Via Anita, Newbury Park, CA 91320 (US).(74) Agent: **MONTGOMERY, Wayne**; Celera Genomics, 45 West Gude Drive, C1-1#316, Rockville, MD 20850 (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, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

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

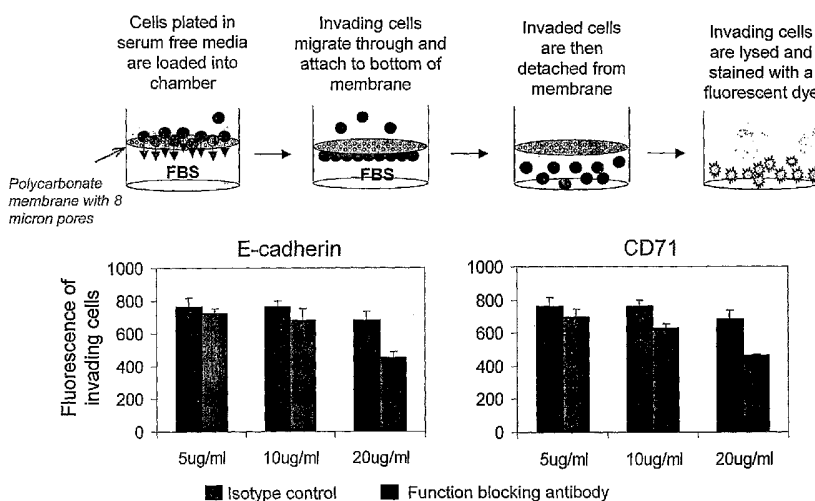
Published:

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

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

(54) Title: METHOD AND COMPOSITIONS FOR TREATING DISEASES TARGETING E-CADHERIN

Effect of Antibodies on Cell Invasion



(57) Abstract: Methods and compositions for diagnosing, detecting and treating a pancreatic disease associated with differential expression of E-cadherin in comparison to healthy cells. Also provided are antagonists or agonists of E-cadherin, and methods for screening agents that modulate the E-cadherin level or activity in vivo or in vitro.

METHOD AND COMPOSITIONS FOR TREATING DISEASES TARGETING E-CADHERIN

FIELD OF THE INVENTION

[0001] This invention relates to the fields of molecular biology and oncology. Specifically, the invention provides a molecular marker and a therapeutic agent for use in the diagnosis and treatment of cancers.

BACKGROUND OF THE INVENTION

[0002] Cancer currently constitutes the second most common cause of death in the United States. Carcinomas of the pancreas are the eighth most prevalent form of cancer and fourth among the most common causes of cancer deaths in this country.

[0003] The prognosis for pancreatic carcinoma is, at present, very poor, it displays the lowest five-year survival rate among all cancers. Such prognosis results primarily from delayed diagnosis, due in part to the fact that the early symptoms are shared with other more common abdominal ailments. Despite the advances in diagnostic imaging methods like ultrasonography (US), endoscopic ultrasonography (EUS), dualphase spiral computer tomography (CT), magnetic resonance imaging (MRT), endoscopic retrograde cholangiopancreatography (ERCP) and transcutaneous or EUS-guided fine-needle aspiration (FNA), distinguishing pancreatic carcinoma from benign pancreatic diseases, especially chronic pancreatitis, is difficult because of the similarities in radiological and imaging features and the lack of specific clinical symptoms for pancreatic carcinoma.

[0004] Substantial efforts have been directed to developing tools useful for early diagnosis of pancreatic carcinomas. Nonetheless, a definitive diagnosis is often dependent on exploratory surgery which is inevitably performed after the disease has advanced past the point when early treatment may be effected.

[0005] One promising method for early diagnosis of various forms of cancer is the identification of specific biochemical moieties, termed targets expressed differentially in the cancerous cells. The targets may be either cell surface proteins or cytosolic proteins. Antibodies or other biomolecules or small molecules which will specifically recognize

and bind to the targets in the cancerous cells potentially provide powerful tools for the diagnosis and treatment of the particular malignancy.

[0006] The epithelial cell adhesion molecule E-cadherin, also known as uvomorulin, mediates calcium-dependent cell adhesion and aggregation. E-cadherin mediated cell-cell contact causes the redistribution of specific membrane proteins, including Na^+ , K^+ -ATPase and fodrin. This redistribution of membrane proteins results in co-localization of the proteins to sites of cell-cell contact. This redistribution of membrane proteins occurs in the absence of tight junctions. Protein complexes containing E-cadherin, ankyrin and fodrin have been identified in kidney cells (McNeill et al., 1990, Cell. 62: 309-316).

[0007] E-cadherins also bind cytoplasmic proteins, including catenins alpha, beta and gamma, through a specific domain in the cytoplasmic region. E-cadherin-catenin complex formation may be dependent upon the phosphorylation of E-cadherin at cytoplasmic locations. The adhesive function of E-cadherin is dependent upon complexing with catenins (Ozawa et al., 1990, Proc. Natl. Acad. Sci. USA. 87: 4246-4250).

[0008] E-cadherin is also expressed in the dorsal root, trigeminal and some other ganglia, possibly playing a role in nerve fiber recognition. Multiple cell types express E-cadherin, including neurons, satellite and Schwann cells (Shimamura et al., 1992, Dev. Biol. 152: 242-254). E-Cadherin, is also called cadherin 1 or type 1.

SUMMARY OF THE INVENTION

[0009] A diseased, e.g. malignant, cell often differs from a normal cell by a differential expression of one or more proteins. These differentially expressed proteins, and suitable fragments thereof, are useful as markers for the diagnosis and treatment of the disease.

[0010] Surprisingly, the present inventors discovered that E-cadherin is differentially expressed in pancreatic tumor cells in comparison to normal pancreatic cells.

Accordingly, the present invention provides methods and compositions for treating pancreatic diseases, especially malignant pancreatic tumors, using E-cadherin as a target.

[0011] In the context of the present invention, the differentially expressed E-cadherin proteins (SEQ ID NOs: 1, 2 and 3) and suitable fragments thereof, and nucleic acids encoding said protein (SEQ ID NOs: 4, 5 and 6, which encode SEQ ID NOs: 1, 2 and 3 respectively) and suitable fragments thereof, are respectfully referred to herein as E-cadherin proteins, E-cadherin peptides or E-cadherin nucleic acids, and collectively as E-cadherin.

[0012] The E-cadherin proteins of the present invention may serve as a target for one or more classes of therapeutic agents, including antibody therapeutics. E-cadherin proteins of the present invention are useful in providing a target for diagnosing a pancreatic cancer or tumor, or predisposition to a pancreatic cancer or tumor mediated by the peptide. Accordingly, the invention provides methods for detecting the presence, or levels of, a E-cadherin protein of the present invention in a biological sample such as tissues, cells and biological fluids isolated from a subject.

[0013] The diagnosis method may detect E-cadherin nucleic acids, proteins, peptides and fragments thereof that are differentially expressed in pancreatic diseases in a test sample, preferably in a biological sample.

[0014] The further embodiment includes but is not limited to, monitoring the disease prognosis (recurrence), diagnosing disease stage, preventing the disease and treating the disease.

[0015] Accordingly, the present invention provides a method for diagnosing or detecting a pancreatic cancer or tumor in a subject comprising: determining the level of E-cadherin in a test sample from said subject, wherein a differential level of said E-cadherin in said sample relative to the level in a control sample from a healthy subject, or the level established for a healthy subject, is indicative of the pancreatic cancer or tumor. The test sample includes but is not limited to a biological sample such as tissue, blood, serum or biological fluid.

[0016] The diagnostic method of the present invention may be suitable for monitoring the disease progression or the treatment progress.

[0017] The diagnostic method of the present invention may be suitable for other epithelial-cell related cancers, such as lung, colon, prostate, ovarian, breast, bladder renal, hepatocellular, pharyngeal, and gastric cancers. The present invention further provides an antagonist to E-cadherin proteins or peptides and a pharmaceutical composition that comprises the antagonist and a suitable carrier. The antagonist may be used for treating the pancreatic disease. Preferably, the antagonist is an antibody that specifically binds to E-cadherin protein or peptide. In another preferred embodiment, the antagonist may be a small molecule that inhibit that function or level of E-cadherin, or an inhibitory nucleic acid molecule, such as an RNAi or antisense molecule against a E-cadherin nucleic acid.

[0018] The present invention provides additionally a pharmaceutical composition comprising an antagonist to E-cadherin of the present invention, and a pharmaceutically acceptable excipient, for treating a pancreatic tumor or cancer.

[0019] The present invention further provides a method for treating pancreatic disease, comprising administering to a patient in need of said treatment a therapeutically effective amount of the pharmaceutical composition.

[0020] The present invention further provides a method for screening for agents that modulate a E-cadherin protein activity, comprising the steps of (i) contacting a candidate agent with a E-cadherin protein, and (ii) assaying for E-cadherin protein activity, wherein a change in said activity in the presence of said agent relative to E-cadherin protein activity in the absence of said agent indicates said agent modulates said E-cadherin protein activity. Candidate agents include but are not limited to protein, peptide, antibody, nucleic acid such as antisense RNA, RNAi fragments, small molecules.

[0021] The screening method may also determine a candidate agent's ability to modulate the expression level of a E-cadherin protein or nucleic acid. The method comprises (i) contacting a candidate agent with a system that is capable of expressing a E-cadherin protein or E-cadherin mRNA, (ii) assaying for the level of a E-cadherin

protein or a E-cadherin mRNA, wherein a specific change in said level in the presence of said agent relative to a level in the absence of said agent indicates said agent modulates said E-cadherin level.

[0022] The present invention further provides a method to screen for agents that bind to the E-cadherin protein, comprising the steps of (i) contacting a test agent with a E-cadherin protein and (ii) measuring the level of binding of agent to said E-cadherin protein.

DESCRIPTION OF FIGURES

[0023] Figure 1. Blocking Effect of E-cadherin antibodies on cell invasion.

[0024] Figure 2. Effect of antibodies on apoptosis of pancreatic cells, wherein E-Cadherin can induce apoptosis in Su86.86 pancreatic cancer cells, however induction of apoptosis was not observed in Hs766t control cells..

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS:

1. E-cadherin Protein and Peptides

[0025] The present invention provides isolated E-cadherin peptide and protein molecules that consisting of, consisting essentially of, or comprising the amino acid sequences of SEQ ID NOs: 1, 2 and 3, respectively encoded by the nucleic acid molecules having the nucleotide sequences of SEQ ID NOs: 4, 5, 6, as well as all obvious variants of these peptides that are within the art to make and use. Some of these variants are described in detail below.

[0026] A E-cadherin peptide or protein can be attached to heterologous sequences to form chimeric or fusion proteins. Such chimeric and fusion proteins comprise a peptide operatively linked to a heterologous protein having an amino acid sequence not

substantially homologous to the peptide. "Operatively linked" indicates that the peptide and the heterologous protein are fused in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the peptide.

[0027] In some uses, the fusion protein does not affect the activity of the peptide or protein per se. For example, the fusion protein can include, but is not limited to, fusion proteins, for example beta-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions, MYC-tagged, HI-tagged and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate the purification of recombinant E-cadherin proteins or peptides. In certain host cells (e.g., mammalian host cells), expression and/or secretion of a protein can be increased by using a heterologous signal sequence.

[0028] A chimeric or fusion E-cadherin protein or peptide can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different protein sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see Ausubel et al., Current Protocols in Molecular Biology, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). E-cadherin-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the E-cadherin protein or peptide.

[0029] Variants of the E-cadherin protein can readily be identified/made using molecular techniques and the sequence information disclosed herein. Further, such variants can readily be distinguished from other peptides based on sequence and/or structural homology to the E-cadherin peptides of the present invention. The degree of homology/identity present will be based primarily on whether the peptide is a functional variant or non-functional variant, the amount of divergence present in the paralog family and the evolutionary distance between the orthologs.

[0030] To determine the percent identity of two amino acid sequences or two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% or more of the length of a reference sequence is aligned for comparison purposes. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

[0031] The comparison of sequences and determination of percent identity and similarity between two sequences can be accomplished using a mathematical algorithm. (Computational Molecular Biology, Lesk, A.M., ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D.W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part 1, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (J. Mol. Biol. (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package, using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (Devereux, J., et al., Nucleic Acids Res. 12(1):387 (1984)), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70,

or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Myers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

[0032] The nucleic acids and protein sequences of the present invention can further be used as a "query sequence" to perform a search against sequence databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (J. Mol. Biol. 215:403-10 (1990)). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (Nucleic Acids Res. 25(17):3389-3402 (1997)). When utilizing BLAST and gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used.

[0033] Allelic variants of a E-cadherin peptide can readily be identified as being a human protein having a high degree (significant) of sequence homology/identity to at least a portion of the E-cadherin peptide as well as being encoded by the same genetic locus as the E-cadherin peptide provided herein. Genetic locus can readily be determined based on the genomic information. As used herein, two proteins (or a region of the proteins) have significant homology when the amino acid sequences are typically at least about 70-80%, 80-90%, and more typically at least about 90-95% or more homologous. A significantly homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid sequence that will hybridize to a E-cadherin peptide encoding nucleic acid molecule under stringent conditions as more fully described below.

[0034] Paralogs of a E-cadherin peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the E-cadherin peptide, as being encoded by a gene from humans, and as having similar activity or function. Two proteins will typically be considered paralogs when the amino acid sequences are typically at least about 60% or greater, and more typically at least about 70% or greater homology through a given region or domain. Such paralogs will be encoded by a nucleic acid sequence that will hybridize to a E-cadherin peptide encoding nucleic acid molecule under moderate to stringent conditions as more fully described below.

[0035] Orthologs of a E-cadherin peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the E-cadherin peptide as well as being encoded by a gene from another organism. Preferred orthologs will be isolated from mammals, preferably primates, for the development of human therapeutic targets and agents. Such orthologs will be encoded by a nucleic acid sequence that will hybridize to a E-cadherin peptide-encoding nucleic acid molecule under moderate to stringent conditions, as more fully described below, depending on the degree of relatedness of the two organisms yielding the proteins.

[0036] Non-naturally occurring variants of the E-cadherin peptides of the present invention can readily be generated using recombinant techniques. Such variants include, but are not limited to deletions, additions and substitutions in the amino acid sequence of the E-cadherin peptide. For example, one class of substitutions is conserved amino acid substitution. Such substitutions are those that substitute a given amino acid in a E-cadherin peptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr; exchange of the acidic residues Asp and Glu; substitution between the amide residues Asn and Gln; exchange of the basic residues Lys and Arg; and replacements among the aromatic residues Phe and Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie et al., Science 247:1306-1310 (1990).

[0037] Variant E-cadherin peptides can be fully functional or can lack function in one or more activities, e.g. ability to bind substrate, ability to phosphorylate substrate, ability to mediate signaling, etc. Fully functional variants typically contain only conservative variation or variation in non-critical residues or in non-critical regions.

[0038] Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncation or a substitution, insertion, inversion, or deletion in a critical residue or critical region.

[0039] Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham et al., Science 244:1081-1085 (1989)). The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as E-cadherin activity or in assays such as an in vitro proliferative activity. Sites that are critical for binding partner/substrate binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith et al., J. Mol. Biol. 224:899-904 (1992); de Vos et al. Science 255:306-312 (1992)).

[0040] The present invention further provides fragments of E-cadherin, in addition to proteins and peptides that comprise and consist of such fragments. As used herein, a fragment comprises at least 8, 10, 12, 14, 16, 18, 20 or more contiguous amino acid residues from E-cadherin. Such fragments can be chosen based on the ability to retain one or more of the biological activities of E-cadherin or could be chosen for the ability to perform a function, e.g. bind a substrate or act as an immunogen. Particularly important fragments are biologically active fragments, peptides that are, for example, about 8 or more amino acids in length. Such fragments will typically comprise a domain or motif of E-cadherin, e.g., active site, a transmembrane domain or a substrate-binding domain. Further, possible fragments include, but are not limited to, domain or motif containing fragments, soluble peptide fragments, and fragments containing immunogenic structures. Predicted domains and functional sites are readily identifiable by computer programs well known and readily available to those of skill in the art (e.g., PROSITE analysis).

[0041] Polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally occurring amino acids. Further, many amino acids, including the terminal amino acids, may be modified by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques well known in the art. Common modifications that occur naturally in E-cadherin are described in basic texts, detailed monographs, and the research literature, and they are well known to those of skill in the art.

[0042] Known modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

[0043] Such modifications are well known to those of skill in the art and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as *Proteins - Structure and Molecular Properties*, 2nd Ed., T.E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as by Wold, F., *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York 1-12 (1983); Seifter et al. (*Meth. Enzymol.* 182: 626-646 (1990)) and Rattan et al. (*Ann. N.Y. Acad. Sci.* 663:48-62 (1992)).

[0044] Accordingly, the E-cadherin of the present invention also encompass derivatives or analogs in which a substituted amino acid residue is not one encoded by

the genetic code, in which a substituent group is included, in which the mature E-cadherin is fused with another compound, such as a compound to increase the half-life of E-cadherin (for example, polyethylene glycol), or in which the additional amino acids are fused to the mature E-cadherin, such as a leader or secretory sequence or a sequence for purification of the mature E-cadherin or a pro-protein sequence.

2. Antibodies against E-cadherin Protein or Fragments Thereof

[0045] Antibodies that selectively bind to the E-cadherin protein or peptides of the present invention can be made using standard procedures known to those of ordinary skills in the art. The term "antibody" is used in the broadest sense, and specifically covers monoclonal antibodies (including full length monoclonal antibodies), polyclonal antibodies, multispecific antibodies (e.g., bispecific antibodies), humanized antibody and antibody fragments (e.g., Fab, F(ab')₂ and Fv) so long as they exhibit the desired biological activity. Antibodies (Abs) and immunoglobulins (Igs) are glycoproteins having the same structural characteristics. While antibodies exhibit binding specificity to a specific antigen, immunoglobulins include both antibodies and other antibody-like molecules that lack antigen specificity.

[0046] As used herein, antibodies are usually heterotetrameric glycoproteins of about 150,000 daltons, composed of two identical light (L) chains and two identical heavy (H) chains. Each light chain is linked to a heavy chain by one covalent disulfide bond, while the number of disulfide linkages varies between the heavy chains of different immunoglobulin isotypes. Each heavy and light chain also has regularly spaced intrachain disulfide bridges. Each heavy chain has at one end a variable domain (VH) followed by a number of constant domains. Each light chain has a variable domain at one end (VL) and a constant domain at its other end. The constant domain of the light chain is aligned with the first constant domain of the heavy chain, and the light chain variable domain is aligned with the variable domain of the heavy chain. Particular amino acid residues are believed to form an interface between the light and heavy chain variable domains. Chothia et al., J. Mol. Biol. 186, 651-63 (1985); Novotny and Haber, Proc. Natl. Acad. Sci. USA 82 4592-4596 (1985).

[0047] An "isolated" antibody is one which has been identified and separated and/or recovered from a component of the environment in which it is produced. Contaminant components of its production environment are materials that would interfere with diagnostic or therapeutic uses for the antibody, and may include enzymes, hormones, and other proteinaceous or nonproteinaceous solutes. In preferred embodiments, the antibody will be purified as measurable by at least three different methods: 1) to greater than 95% by weight of antibody as determined by the Lowry method, and most preferably more than 99% by weight; 2) to a degree sufficient to obtain at least 15 residues of N-terminal or internal amino acid sequence by use of a spinning cup sequenator; or 3) to homogeneity by SDS-PAGE under reducing or non-reducing conditions using Coomassie blue or, preferably, silver stain. Isolated antibody includes the antibody in situ within recombinant cells since at least one component of the antibody's natural environment will not be present. Ordinarily, however, isolated antibody will be prepared by at least one purification step.

[0048] An "antigenic region" or "antigenic determinant" or an "epitope" includes any protein determinant capable of specific binding to an antibody. This is the site on an antigen to which each distinct antibody molecule binds. Epitopic determinants usually consist of active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three-dimensional structural characteristics, as well as charge characteristics.

[0049] "Antibody specificity," is an antibody, which has a stronger binding affinity for an antigen from a first subject species than it has for a homologue of that antigen from a second subject species. Normally, the antibody "bind specifically" to a human antigen (i.e., has a binding affinity (Kd) value of no more than about 1×10^{-7} M, preferably no more than about 1×10^{-8} M and most preferably no more than about 1×10^{-9} M) but has a binding affinity for a homologue of the antigen from a second subject species which is at least about 50 fold, or at least about 500 fold, or at least about 1000 fold, weaker than its binding affinity for the human antigen. The antibody can be of any of the various types of antibodies as defined above, but preferably is a humanized or human antibody (Queen et al., US Patent Nos. 5,530,101; 5,585,089; 5,693,762; and 6,180,370).

[0050] The present invention provides an “antibody variant,” which refers to an amino acid sequence variant of an antibody wherein one or more of the amino acid residues have been modified. Such variant necessarily have less than 100% sequence identity or similarity with the amino acid sequence having at least 75% amino acid sequence identity or similarity with the amino acid sequence of either the heavy or light chain variable domain of the antibody, more preferably at least 80%, more preferably at least 85%, more preferably at least 90%, and most preferably at least 95%. Since the method of the invention applies equally to both polypeptides, antibodies and fragments thereof, these terms are sometimes employed interchangeably.

[0051] The term “antibody fragment” refers to a portion of a full-length antibody, generally the antigen binding or variable region. Examples of antibody fragments include Fab, Fab', F(ab')₂ and Fv fragments. Papain digestion of antibodies produces two identical antigen binding fragments, called the Fab fragment, each with a single antigen binding site, and a residual “Fc” fragment, so-called for its ability to crystallize readily. Pepsin treatment yields an F(ab')₂ fragment that has two antigen binding fragments which are capable of crosslinking antigen, and a residual other fragment (which is termed pFc'). Additional fragments can include diabodies, linear antibodies, single-chain antibody molecules, and multispecific antibodies formed from antibody fragments. As used herein, “functional fragment” with respect to antibodies, refers to Fv, F(ab) and F(ab')₂ fragments.

[0052] An “Fv” fragment is the minimum antibody fragment that contains a complete antigen recognition and binding site. This region consists of a dimer of one heavy and one light chain variable domain in a tight, non-covalent association (V_H-V_L dimer). It is in this configuration that the three CDRs of each variable domain interact to define an antigen-binding site on the surface of the V_H-V_L dimer. Collectively, the six CDRs confer antigen-binding specificity to the antibody. However, even a single variable domain (or half of an Fv comprising only three CDRs specific for an antigen) has the ability to recognize and bind antigen, although at a lower affinity than the entire binding site.

[0053] The Fab fragment [also designated as F(ab)] also contains the constant domain of the light chain and the first constant domain (CH1) of the heavy chain. Fab' fragments differ from Fab fragments by the addition of a few residues at the carboxyl terminus of the heavy chain CH1 domain including one or more cysteines from the antibody hinge region. Fab'-SH is the designation herein for Fab' in which the cysteine residue(s) of the constant domains have a free thiol group. F(ab') fragments are produced by cleavage of the disulfide bond at the hinge cysteines of the F(ab')₂ pepsin digestion product. Additional chemical couplings of antibody fragments are known to those of ordinary skill in the art.

[0054] The present invention further provides monoclonal antibody, polyclonal antibody as well as humanized antibody. In general, to generate antibodies, an isolated peptide is used as an immunogen and is administered to a mammalian organism, such as a rat, rabbit or mouse. The full-length protein, an antigenic peptide fragment or a fusion protein of the E-cadherin protein can be used. Particularly important fragments are those covering functional domains. Many methods are known for generating and/or identifying antibodies to a given target peptide. Several such methods are described by Harlow, *Antibodies*, Cold Spring Harbor Press, (1989).

[0055] The term "monoclonal antibody" as used herein refers to an antibody obtained from a population of substantially homogeneous antibodies, i.e., the individual antibodies comprising the population are identical except for possible naturally occurring mutations that may be present in minor amounts. Monoclonal antibodies are highly specific, being directed against a single antigenic site. Furthermore, in contrast to conventional (polyclonal) antibody preparations which typically include different antibodies directed against different determinants (epitopes), each monoclonal antibody is directed against a single determinant on the antigen. In addition to their specificity, the monoclonal antibodies are advantageous in that they are synthesized by the hybridoma culture, uncontaminated by other immunoglobulins. The modifier "monoclonal" antibody indicates the character of the antibody as being obtained from a substantially homogeneous population of antibodies, and is not to be construed as requiring production of the antibody by any particular method. For example, the monoclonal antibodies to be

used in accordance with the present invention may be made by the hybridoma method first described by Kohler and Milstein, *Nature* 256, 495 (1975), or may be made by recombinant methods, e.g., as described in U.S. Pat. No. 4,816,567. The monoclonal antibodies for use with the present invention may also be isolated from phage antibody libraries using the techniques described in Clackson et al., *Nature* 352: 624-628 (1991), as well as in Marks et al., *J. Mol. Biol.* 222: 581-597 (1991). For detailed procedures for making a monoclonal antibody, see the Example below.

[0056] "Humanized" forms of non-human (e.g. murine or rabbit) antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', F(ab')₂ or other antigen-binding subsequences of antibodies) which contain minimal sequence derived from non-human immunoglobulin. For the most part, humanized antibodies are human immunoglobulins (recipient antibody) in which residues from a complementary determining region (CDR) of the recipient are replaced by residues from a CDR of a non-human species (donor antibody) such as mouse, rat or rabbit having the desired specificity, affinity and capacity. In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Furthermore, humanized antibody may comprise residues, which are found neither in the recipient antibody nor in the imported CDR or framework sequences. These modifications are made to further refine and optimize antibody performance. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the FR regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin. For further details, see: Jones et al., *Nature* 321, 522-525 (1986); Reichmann et al., *Nature* 332, 323-327 (1988) and Presta, *Curr. Op. Struct. Biol.* 2, 593-596 (1992).

[0057] Polyclonal antibodies may be prepared by any known method or modifications of these methods including obtaining antibodies from patients. For example, a complex of an immunogen such as E-cadherin protein, peptides or fragments

thereof and a carrier protein is prepared and an animal is immunized by the complex according to the same manner as that described with respect to the above monoclonal antibody preparation and the description in the Example. A serum or plasma containing the antibody against the protein is recovered from the immunized animal and the antibody is separated and purified. The gamma globulin fraction or the IgG antibodies can be obtained, for example, by use of saturated ammonium sulfate or DEAE SEPHADEX, or other techniques known to those skilled in the art.

[0058] The antibody titer in the antiserum can be measured according to the same manner as that described above with respect to the supernatant of the hybridoma culture. Separation and purification of the antibody can be carried out according to the same separation and purification method of antibody as that described with respect to the above monoclonal antibody and in the Example.

[0059] The protein used herein as the immunogen is not limited to any particular type of immunogen. In one aspect, antibodies are preferably prepared from regions or discrete fragments of the E-cadherin protein. Antibodies can be prepared from any region of the peptide as described herein. In particular, they are selected from a group consisting of SEQ ID NOS: 1-3 and fragments thereof. An antigenic fragment will typically comprise at least 8 contiguous amino acid residues. The antigenic peptide can comprise, however, at least 10, 12, 14, 16 or more amino acid residues. Such fragments can be selected on a physical property, such as fragments correspond to regions that are located on the surface of the protein, e.g., hydrophilic regions or can be selected based on sequence uniqueness.

[0060] Antibodies may also be produced by inducing production in the lymphocyte population or by screening antibody libraries or panels of highly specific binding reagents as disclosed in Orlandi et al. (1989; Proc Natl Acad Sci 86:3833-3837) or Winter et al. (1991; Nature 349:293-299). A protein may be used in screening assays of phagemid or B-lymphocyte immunoglobulin libraries to identify antibodies having a desired specificity. Numerous protocols for competitive binding or immunoassays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Smith G. P., 1991, Curr. Opin. Biotechnol. 2: 668-673.

[0061] The antibodies of the present invention can also be generated using various phage display methods known in the art. In phage display methods, functional antibody domains are displayed on the surface of phage particles which carry the polynucleotide sequences encoding them. In a particular, such phage can be utilized to display antigen-binding domains expressed from a repertoire or combinatorial antibody library (e.g., human or murine). Phage expressing an antigen binding domain that binds the antigen of interest can be selected or identified with antigen, e.g., using labeled antigen or antigen bound or captured to a solid surface or bead. Phage used in these methods are typically filamentous phage including fd and M13 binding domains expressed from phage with Fab, Fv or disulfide stabilized Fv antibody domains recombinantly fused to either the phage gene III or gene VIII protein. Examples of phage display methods that can be used to make the antibodies of the present invention include those disclosed in Brinkman et al., J. Immunol. Methods 182:41-50 (1995); Ames et al., J. Immunol. Methods 184:177-186 (1995); Kettleborough et al., Eur. J. Immunol. 24:952-958 (1994); Persic et al., Gene, 187: 9-18 (1997); Burton et al., Advances in Immunology 57:191-280 (1994); PCT application No. PCT/GB91/01134; PCT publications WO 90/02809; WO 91/10737; WO 92/01047; WO 92/18619; WO 93/11236; WO 95/15982; WO 95/20401; and U.S. Pat. Nos. 5,698,426; 5,223,409; 5,403,484; 5,580,717; 5,427,908; 5,750,753; 5,821,047; 5,571,698; 5,427,908; 5,516,637; 5,780,225; 5,658,727; 5,733,743 and 5,969,108; each of which is incorporated herein by reference in its entirety.

[0062] Antibody can be also made recombinantly. When using recombinant techniques, the antibody variant can be produced intracellularly, in the periplasmic space, or directly secreted into the medium. If the antibody variant is produced intracellularly, as a first step, the particulate debris, either host cells or lysed fragments, is removed, for example, by centrifugation or ultrafiltration. Carter et al., Bio/Technology 10: 163-167 (1992) describe a procedure for isolating antibodies that are secreted to the periplasmic space of *E. coli*. Briefly, cell paste is thawed in the presence of sodium acetate (pH 3.5), EDTA, and phenylmethylsulfonylfluoride (PMSF) over about 30 minutes. Cell debris can be removed by centrifugation. Where the antibody variant is secreted into the medium, supernatants from such expression systems are generally first concentrated using a commercially available protein concentration filter, for example, an Amicon or Millipore

PELLICON ultrafiltration unit. A protease inhibitor such as PMSF may be included in any of the foregoing steps to inhibit proteolysis and antibiotics may be included to prevent the growth of adventitious contaminants.

[0063] The antibodies or antigen binding fragments may also be produced by genetic engineering. The technology for expression of both heavy and light chain genes in *E. coli* is the subject the following PCT patent applications; publication number WO 901443, WO901443, and WO 9014424 and in Huse et al., 1989 Science 246:1275-1281. The general recombinant methods are well known in the art.

[0064] The antibody composition prepared from the cells can be purified using, for example, hydroxylapatite chromatography, gel electrophoresis, dialysis, and affinity chromatography, with affinity chromatography being the preferred purification technique. The suitability of protein A as an affinity ligand depends on the species and isotype of any immunoglobulin Fc domain that is present in the antibody. Protein A can be used to purify antibodies that are based on human .delta.1, .delta.2 or .delta.4 heavy chains (Lindmark et al., J. Immunol Meth. 62: 1-13 (1983)). Protein G is recommended for all mouse isotypes and for human .delta.3 (Guss et al., EMBO J. 5: 1567-1575 (1986)). The matrix to which the affinity ligand is attached is most often agarose, but other matrices are available. Mechanically stable matrices such as controlled pore glass or poly(styrenedivinyl)benzene allow for faster flow rates and shorter processing times than can be achieved with agarose. Where the antibody comprises a CH3 domain, the BAKERBOND ABXTM resin (J.T. Baker, Phillipsburg, N.J.) is useful for purification. Other techniques for protein purification such as fractionation on an ion-exchange column, ethanol precipitation, Reverse Phase HPLC, chromatography on silica, chromatography on heparin SEPHAROSE chromatography on an anion or cation exchange resin (such as a polyaspartic acid column), chromatofocusing, SDS-PAGE, and ammonium sulfate precipitation are also available depending on the antibody to be recovered.

[0065] Following any preliminary purification step(s), the mixture comprising the antibody of interest and contaminants may be subjected to low pH hydrophobic

interaction chromatography using an elution buffer at a pH between about 2.5-4.5, preferably performed at low salt concentrations (e.g., from about 0-0.25M salt).

3. E-cadherin Nucleic Acid Molecules

[0066] Isolated E-cadherin nucleic acid molecules of the present invention consist of, consist essentially of, or comprise a nucleotide sequence that encodes E-cadherin peptides of the present invention, an allelic variant thereof, or an ortholog or paralog thereof. As used herein, an "isolated" nucleic acid molecule is one that is separated from other nucleic acid present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. However, there can be some flanking nucleotide sequences, for example up to about 5KB, 4KB, 3KB, 2KB, or 1KB or less, particularly contiguous peptide encoding sequences and peptide encoding sequences within the same gene but separated by introns in the genomic sequence. The important point is that the nucleic acid is isolated from remote and unimportant flanking sequences such that it can be subjected to the specific manipulations described herein such as recombinant expression, preparation of probes and primers, and other uses specific to the nucleic acid sequences.

[0067] Moreover, an "isolated" nucleic acid molecule, such as a transcript/cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. However, the nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated.

[0068] For example, recombinant DNA molecules contained in a vector are considered isolated. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include in vivo or in vitro RNA transcripts of the isolated DNA molecules of the present invention. Isolated

nucleic acid molecules according to the present invention further include such molecules produced synthetically.

[0069] The isolated nucleic acid molecules can encode the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature peptide (when the mature form has more than one peptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, facilitate protein trafficking, prolong or shorten protein half-life or facilitate manipulation of a protein for assay or production, among other things. As generally is the case in situ, the additional amino acids may be processed away from the mature protein by cellular enzymes.

[0070] As mentioned above, the isolated nucleic acid molecules include, but are not limited to, the sequence encoding E-cadherin peptide alone, the sequence encoding the mature peptide and additional coding sequences, such as a leader or secretory sequence (e.g., a pre-pro or pro-protein sequence), the sequence encoding the mature peptide, with or without the additional coding sequences, plus additional non-coding sequences, for example introns and non-coding 5' and 3' sequences such as transcribed but non-translated sequences that play a role in transcription, mRNA processing (including splicing and polyadenylation signals), ribosome binding and stability of mRNA. In addition, the nucleic acid molecule may be fused to a marker sequence encoding, for example, a peptide that facilitates purification.

[0071] Isolated nucleic acid molecules can be in the form of RNA, such as mRNA, or in the form DNA, including cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The nucleic acid, especially DNA, can be double-stranded or single-stranded. Single-stranded nucleic acid can be the coding strand (sense strand) or the non-coding strand (anti-sense strand).

[0072] The invention further provides nucleic acid molecules that encode fragments of the proteins of the present invention as well as nucleic acid molecules that encode obvious variants of E-cadherin protein of the present invention that are described above. Such nucleic acid molecules may be naturally occurring, such as allelic variants (same

locus), paralogs (different locus), and orthologs (different organism), or may be constructed by recombinant DNA methods or by chemical synthesis. Such non-naturally occurring variants may be made by mutagenesis techniques, including those applied to nucleic acid molecules, cells, or organisms. Accordingly, as discussed above, the variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions.

[0073] A fragment comprises a contiguous nucleotide sequence greater than 12 or more nucleotides. Further, a fragment could at least 30, 40, 50, 100, 250 or 500 nucleotides in length. The length of the fragment will be based on its intended use. For example, the fragment can encode epitope bearing regions of the peptide, or can be useful as DNA probes and primers. Such fragments can be isolated using the known nucleotide sequence to synthesize an oligonucleotide probe. A labeled probe can then be used to screen a cDNA library, genomic DNA library, or mRNA to isolate nucleic acid corresponding to the coding region. Further, primers can be used in PCR reactions to clone specific regions of gene.

[0074] A probe/primer typically comprises substantially a purified oligonucleotide or oligonucleotide pair. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 20, 25, 40, 50 or more consecutive nucleotides.

[0075] Orthologs, homologs, and allelic variants can be identified using methods well known in the art. As described in the Peptide Section, these variants comprise a nucleotide sequence encoding a peptide that is typically 60-70%, 70-80%, 80-90%, and more typically at least about 90-95% or more homologous to the nucleotide sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under moderate to stringent conditions, to the nucleotide sequence shown in the Figure sheets or a fragment of the sequence. Allelic variants can readily be determined by genetic locus of the encoding gene.

[0076] As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences encoding a peptide at least 60-70% homologous to each other typically remain hybridized to each other. The conditions can be such that sequences at least about 60%, at least about 70%, or at least about 80% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. One example of stringent hybridization conditions is hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C. Examples of moderate to low stringency hybridization conditions are well known in the art.

4. Vectors and Host Cells

[0077] The invention also provides vectors containing the nucleic acid molecules described herein. The term "vector" refers to a vehicle, preferably a nucleic acid molecule, which can transport the nucleic acid molecules. When the vector is a nucleic acid molecule, the nucleic acid molecules are covalently linked to the vector nucleic acid. With this aspect of the invention, the vector includes a plasmid, single or double stranded phage, a single or double stranded RNA or DNA viral vector, or artificial chromosome, such as a BAC, PAC, YAC, OR MAC.

[0078] A vector can be maintained in the host cell as an extrachromosomal element where it replicates and produces additional copies of the nucleic acid molecules. Alternatively, the vector may integrate into the host cell genome and produce additional copies of the nucleic acid molecules when the host cell replicates.

[0079] The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of the nucleic acid molecules. The vectors can function in prokaryotic or eukaryotic cells or in both (shuttle vectors).

[0080] Expression vectors contain cis-acting regulatory regions that are operably linked in the vector to the nucleic acid molecules such that transcription of the nucleic

acid molecules is allowed in a host cell. The nucleic acid molecules can be introduced into the host cell with a separate nucleic acid molecule capable of affecting transcription. Thus, the second nucleic acid molecule may provide a trans-acting factor interacting with the cis-regulatory control region to allow transcription of the nucleic acid molecules from the vector. Alternatively, a trans-acting factor may be supplied by the host cell. Finally, a trans-acting factor can be produced from the vector itself. It is understood, however, that in some embodiments, transcription and/or translation of the nucleic acid molecules can occur in a cell-free system.

[0081] The regulatory sequences to which the nucleic acid molecules described herein can be operably linked include promoters for directing mRNA transcription. These include, but are not limited to, the left promoter from bacteriophage, the lac, TRP, and TAC promoters from *E. coli*, the early and late promoters from SV40, the CMV immediate early promoter, the adenovirus early and late promoters, and retrovirus long-terminal repeats.

[0082] In addition to control regions that promote transcription, expression vectors may also include regions that modulate transcription, such as repressor binding sites and enhancers. Examples include the SV40 enhancer, the cytomegalovirus immediate early enhancer, polyoma enhancer, adenovirus enhancers, and retrovirus LTR enhancers.

[0083] In addition to containing sites for transcription initiation and control, expression vectors can also contain sequences necessary for transcription termination and, in the transcribed region a ribosome binding site for translation. Other regulatory control elements for expression include initiation and termination codons as well as polyadenylation signals. The person of ordinary skill in the art would be aware of the numerous regulatory sequences that are useful in expression vectors. Such regulatory sequences are described, for example, in Sambrook et al., *Molecular Cloning: A Laboratory Manual*. 3rd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (2001).

[0084] A variety of expression vectors can be used to express a nucleic acid molecule. Such vectors include chromosomal, episomal, and virus-derived vectors, for

example vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, including yeast artificial chromosomes, from viruses such as baculoviruses, papovaviruses such as SV40, Vaccinia viruses, adenoviruses, poxviruses, pseudorabies viruses, and retroviruses. Vectors may also be derived from combinations of these sources such as those derived from plasmid and bacteriophage genetic elements, e.g. cosmids and phagemids. Appropriate cloning and expression vectors for prokaryotic and eukaryotic hosts are described in Sambrook et al., *Molecular Cloning: A Laboratory Manual*. 3rd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (2001).

[0085] The regulatory sequence may provide constitutive expression in one or more host cells (i.e. tissue specific) or may provide for inducible expression in one or more cell types such as by temperature, nutrient additive, or exogenous factor such as a hormone or other ligand. A variety of vectors providing for constitutive and inducible expression in prokaryotic and eukaryotic hosts are well known to those of ordinary skill in the art.

[0086] The nucleic acid molecules can be inserted into the vector nucleic acid by well-known methodology. Generally, the DNA sequence that will ultimately be expressed is joined to an expression vector by cleaving the DNA sequence and the expression vector with one or more restriction enzymes and then ligating the fragments together. Procedures for restriction enzyme digestion and ligation are well known to those of ordinary skill in the art.

[0087] The vector containing the appropriate nucleic acid molecule can be introduced into an appropriate host cell for propagation or expression using well-known techniques. Bacterial cells include, but are not limited to, *E. coli*, *Streptomyces*, and *Salmonella typhimurium*. Eukaryotic cells include, but are not limited to, yeast, insect cells such as *Drosophila*, animal cells such as COS and CHO cells, and plant cells.

[0088] As described herein, it may be desirable to express the peptide as a fusion protein. Accordingly, the invention provides fusion vectors that allow for the production of the peptides. Fusion vectors can increase the expression of a recombinant protein; increase the solubility of the recombinant protein, and aid in the purification of the

protein by acting for example as a ligand for affinity purification. A proteolytic cleavage site may be introduced at the junction of the fusion moiety so that the desired peptide can ultimately be separated from the fusion moiety. Proteolytic enzymes include, but are not limited to, factor Xa, thrombin, and enterokinase. Typical fusion expression vectors include pGEX (Smith et al., Gene 67:31-40 (1988)), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al., Gene 69:301-315 (1988)) and pET 11d (Studier et al., Gene Expression Technology: Methods in Enzymology 185:60-89 (1990)).

[0089] Recombinant protein expression can be maximized in host bacteria by providing a genetic background wherein the host cell has an impaired capacity to proteolytically cleave the recombinant protein. (Gottesman, S., Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, California (1990) 119-128). Alternatively, the sequence of the nucleic acid molecule of interest can be altered to provide preferential codon usage for a specific host cell, for example *E. coli*. (Wada et al., Nucleic Acids Res. 20:2111-2118 (1992)).

[0090] The nucleic acid molecules can also be expressed by expression vectors suitable in a yeast host. Examples of vectors for expression in yeast e.g., *S. cerevisiae* include pYepSec1 (Baldari, et al., EMBO J. 6:229-234 (1987)), pMFa (Kurjan et al., Cell 30:933-943(1982)), pJRY88 (Schultz et al., Gene 54:113-123 (1987)), and pYES2 (Invitrogen Corporation, San Diego, CA).

[0091] The nucleic acid molecules can also be expressed in insect cells using, for example, baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith et al., Mol. Cell Biol. 3:2156-2165 (1983)) and the pVL series (Lucklow et al., Virology 170:31-39 (1989)).

[0092] In certain embodiments of the invention, the nucleic acid molecules described herein are expressed in mammalian cells using mammalian expression vectors. Examples

of mammalian expression vectors include pCDM8 (Seed, B. Nature 329:840(1987)) and pMT2PC (Kaufman et al., EMBO J. 6:187-195 (1987)).

[0093] The expression vectors listed herein are provided by way of example only of the well-known vectors available to those of ordinary skill in the art that would be useful to express the nucleic acid molecules. The person of ordinary skill in the art would be aware of other vectors suitable for maintenance propagation or expression of the nucleic acid molecules described herein. These are found for example in Sambrook, J., Fritsh, E. F., and Maniatis, T. Molecular Cloning: A Laboratory Manual. 3rd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (2001).

[0094] The invention also encompasses vectors in which the nucleic acid sequences described herein are cloned into the vector in reverse orientation, but operably linked to a regulatory sequence that permits transcription of antisense RNA. Thus, an antisense transcript can be produced to all, or to a portion, of the nucleic acid molecule sequences described herein, including both coding and non-coding regions. Expression of this antisense RNA is subject to each of the parameters described above in relation to expression of the sense RNA (regulatory sequences, constitutive or inducible expression, tissue-specific expression).

[0095] The invention also relates to recombinant host cells containing the vectors described herein. Host cells therefore include prokaryotic cells, lower eukaryotic cells such as yeast, other eukaryotic cells such as insect cells, and higher eukaryotic cells such as mammalian cells.

[0096] The recombinant host cells are prepared by introducing the vector constructs described herein into the cells by techniques readily available to the person of ordinary skill in the art. These include, but are not limited to, calcium phosphate transfection, DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, lipofection, and other techniques such as those found in Sambrook, et al. (Molecular Cloning: A Laboratory Manual. 3rd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (2001)).

[0097] Host cells can contain more than one vector. Thus, different nucleotide sequences can be introduced on different vectors of the same cell. Similarly, the nucleic acid molecules can be introduced either alone or with other nucleic acid molecules that are not related to the nucleic acid molecules such as those providing trans-acting factors for expression vectors. When more than one vector is introduced into a cell, the vectors can be introduced independently, co-introduced or joined to the nucleic acid molecule vector.

[0098] In the case of bacteriophage and viral vectors, these can be introduced into cells as packaged or encapsulated virus by standard procedures for infection and transduction. Viral vectors can be replication-competent or replication-defective. In the case in which viral replication is defective, replication will occur in host cells providing functions that complement the defects.

[0099] Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the recombinant vector constructs. The marker can be contained in the same vector that contains the nucleic acid molecules described herein or may be on a separate vector. Markers include tetracycline or ampicillin-resistance genes for prokaryotic host cells and dihydrofolate reductase or neomycin resistance for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait will be effective.

[00100] While the mature proteins can be produced in bacteria, yeast, mammalian cells, and other cells under the control of the appropriate regulatory sequences, cell-free transcription and translation systems can also be used to produce these proteins using RNA derived from the DNA constructs described herein.

[00101] Where secretion of the peptide is desired, which may be difficult to achieve with a multi-transmembrane domain containing protein such as E-cadherin, appropriate secretion signals are incorporated into the vector. The signal sequence can be endogenous to the peptides or heterologous to these peptides.

[00102] Where the peptide is not secreted into the medium, the protein can be isolated from the host cell by standard disruption procedures, including freeze thaw, sonication, mechanical disruption, use of lysing agents and the like. The peptide can then be recovered and purified by well-known purification methods including ammonium sulfate precipitation, acid extraction, anion or cationic exchange chromatography, phosphocellulose chromatography, hydrophobic-interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography, or high performance liquid chromatography.

[00103] It is also understood that depending upon the host cell in recombinant production of the peptides described herein, the peptides can have various glycosylation patterns, depending upon the cell, or maybe non-glycosylated as when produced in bacteria. In addition, the peptides may include an initial modified methionine in some cases as a result of a host-mediated process.

[00104] The recombinant host cells expressing the peptides described herein have a variety of uses. First, the cells are useful for producing E-cadherin protein or peptide that can be further purified to produce desired amounts of E-cadherin protein or fragments. Thus, host cells containing expression vectors are useful for peptide production.

[00105] Host cells are also useful for conducting cell-based assays involving the E-cadherin protein or E-cadherin protein fragments, such as those described above as well as other formats known in the art. Thus, a recombinant host cell expressing a native E-cadherin protein is useful for assaying compounds that stimulate or inhibit E-cadherin protein function.

[00106] Host cells are also useful for identifying E-cadherin protein mutants in which these functions are affected. If the mutants naturally occur and give rise to a pathology, host cells containing the mutations are useful to assay compounds that have a desired effect on the mutant E-cadherin protein (for example, stimulating or inhibiting function) which may not be indicated by their effect on the native E-cadherin protein.

5. Detection and Diagnosis in General

[00107] As used herein, a “biological sample” can be collected from tissues, blood, sera, cell lines or biological fluids such as, plasma, interstitial fluid, urine, cerebrospinal fluid, and the like, containing cells. In preferred embodiments, a biological sample comprises cells or tissues suspected of having diseases (e.g., cells obtained from a biopsy).

[00108] As used herein, a “differential level” is defined as the level of E-cadherin protein or nucleic acids in a test sample either above or below the level in control samples, wherein the level of control samples is obtained either from a control cell line, a normal tissue or body fluids, or combination thereof, from a healthy subject. While the protein is overexpressed, the expression of E-cadherin is preferably greater than about 20%, or preferably greater than about 30%, and most preferably greater than about 50% or more of pancreatic disease sample, at a level that is at least two fold, and preferably at least five fold, greater than the level of expression in control samples, as determined using a representative assay provided herein. While the protein is under expressed, the expression of E-cadherin is preferably less than about 20%, or preferably less than 30%, and most preferably less than about 50% or more of the pancreatic disease sample, at a level that is at least 0.5 fold, and preferably at least 0.2 fold less than the level of the expression in control samples, as determined using a representative assay provided herein.

[00109] As used herein, a “subject” can be a mammalian subject or non mammalian subject, preferably, a mammalian subject. A mammalian subject can be human or non-human, preferably human. A healthy subject is defined as a subject without detectable pancreatic diseases or pancreatic associated diseases by using conventional diagnostic methods.

[00110] As used herein, the “disease(s)” include pancreatic diseases and pancreatic associated disease. Preferably, the disease is a pancreatic cancer.

6. Treatment in General

[00111] This invention further pertains to novel agents identified by the screening assays described below. It is also within the scope of this invention to use an agent identified for treatment purposes. For example, an agent identified as described herein (e.g., a E-cadherin-modulating agent, an antisense E-cadherin nucleic acid molecule, an E-cadherin-RNAi fragment, a E-cadherin-specific antibody, or a E-cadherin-binding partner) can be used in an animal or other model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal or other model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

[00112] Modulators of E-cadherin protein activity identified according to these drug screening assays can be used to treat a subject with a disorder mediated by E-cadherin, e.g. by treating cells or tissues that express E-cadherin at a differential level. Methods of treatment include the steps of administering a modulator of E-cadherin activity in a pharmaceutical composition to a subject in need of such treatment.

[00113] The following terms, as used in the present specification and claims, are intended to have the meaning as defined below, unless indicated otherwise.

[00114] "Treat," "treating" or "treatment" of a disease includes: (1) inhibiting the disease, i.e., arresting or reducing the development of the disease or its clinical symptoms, or (2) relieving the disease, i.e., causing regression of the disease or its clinical symptoms.

[00115] The term "prophylaxis" is used to distinguish from "treatment," and to encompass both "preventing" and "suppressing," it is not always possible to distinguish between "preventing" and "suppressing," as the ultimate inductive event or events may be unknown, latent, or the patient is not ascertained until well after the occurrence of the event or events. Therefore, the term "protection," as used herein, is meant to include "prophylaxis."

[00116] A “therapeutically effective amount” means the amount of an agent that, when administered to a subject for treating a disease, is sufficient to effect such treatment for the disease. The “therapeutically effective amount” will vary depending on the agent, the disease and its severity and the age, weight, etc., of the subject to be treated.

[00117] A “pancreatic disease” includes pancreatic cancer, pancreatic tumor (exocrine or endocrine), pancreatic cysts, acute pancreatitis, chronic pancreatitis, diabetes (type I and II) as well as pancreatic trauma. The method of the present invention is preferably used for treating a pancreatic cancer.

[00118] In one embodiment, when decreased expression or activity of the protein is desired, an inhibitor, antagonist, antibody and the like or a pharmaceutical agent containing one or more of these molecules may be delivered. Such delivery may be effected by methods well known in the art and may include delivery by an antibody specifically targeted to the protein.

[00119] In another embodiment, when increased expression or activity of the protein is desired, the protein, an agonist, an enhancer and the like or a pharmaceutical agent containing one or more of these molecules may be delivered. Such delivery may be effected by methods well known in the art.

[00120] While it is possible for the modulating agent to be administered in a pure or substantially pure form, it is preferable to present it as a pharmaceutical composition, formulation or preparation with a carrier. The formulations of the present invention, both for veterinary and for human use, comprise a suitable active E-cadherin modulating agent, together with one or more pharmaceutically acceptable carriers and, optionally, other therapeutic ingredients. The carrier(s) must be “acceptable” in the sense of being compatible with the other ingredients of the formulation and not deleterious to the recipient thereof. The formulations may conveniently be presented in unit dosage form and may be prepared by any method well-known in the pharmaceutical art.

[00121] Suitable pharmaceutical carriers include proteins such as albumins (e.g., U.S. Pat. No. 4,507,234, to Kato et al.), peptides and polysaccharides such as aminodextran

(e.g., U.S. Pat. No. 4,699,784, to Shih et al.), or water. A carrier may also bear an agent by noncovalent bonding or by encapsulation, such as within a liposome vesicle (e.g., U.S. Pat. Nos. 4,429,008 and 4,873,088). Carriers specific for radionuclide agents include radiohalogenated small molecules and chelating compounds. For example, U.S. Pat. No. 4,735,792 discloses representative radiohalogenated small molecules and their synthesis. A radionuclide chelate may be formed from chelating compounds that include those containing nitrogen and sulfur atoms as the donor atoms for binding the metal, metal oxide, radionuclide. For example, U.S. Pat. No. 4,673,562, to Davison et al. discloses representative chelating compounds and their synthesis.

[00122] All methods include the step of bringing into association the active ingredient with the carrier, which constitutes one or more accessory ingredients. In general, the formulations are prepared by uniformly and intimately bringing into association the active ingredient with liquid carriers or finely divided solid carriers or both, and then, if necessary, shaping the product into the desired formulation.

[00123] Formulations suitable for intravenous intramuscular, subcutaneous, or intraperitoneal administration conveniently comprise sterile aqueous solutions of the active ingredient with solutions, which are preferably isotonic with the blood of the recipient. Such formulations may be conveniently prepared by dissolving solid active ingredient in water containing physiologically compatible substances such as sodium chloride (e.g. 0.1-2.0M), glycine, and the like, and having a buffered pH compatible with physiological conditions to produce an aqueous solution, and rendering said solution sterile. These may be present in unit or multi-dose containers, for example, sealed ampoules or vials.

[00124] The formulations of the present invention may incorporate a stabilizer. Illustrative stabilizers are polyethylene glycol, proteins, saccharides, amino acids, inorganic acids, and organic acids, which may be used either on their own or as admixtures. These stabilizers are preferably incorporated in an amount of 0.11-10,000 parts by weight per part by weight of immunogen. If two or more stabilizers are to be used, their total amount is preferably within the range specified above. These stabilizers

are used in aqueous solutions at the appropriate concentration and pH. The specific osmotic pressure of such aqueous solutions is generally in the range of 0.1-3.0 osmoles, preferably in the range of 0.8-1.2. The pH of the aqueous solution is adjusted to be within the range of 5.0-9.0, preferably within the range of 6-8. In formulating the antibody of the present invention, anti-adsorption agent may be used.

[00125] Additional pharmaceutical methods may be employed to control the duration of action. Controlled release preparations may be achieved through the use of polymer to complex or absorb the proteins or their derivatives. The controlled delivery may be exercised by selecting appropriate macromolecules (for example polyester, polyamino acids, polyvinyl, pyrrolidone, ethylenevinylacetate, methylcellulose, carboxymethylcellulose, or protamine sulfate) and the concentration of macromolecules as well as the methods of incorporation in order to control release. Another possible method to control the duration of action by controlled-release preparations is to incorporate anti- E-cadherin antibody into particles of a polymeric material such as polyesters, polyamino acids, hydrogels, poly(lactic acid) or ethylene vinylacetate copolymers. Alternatively, instead of incorporating these agents into polymeric particles, it is possible to entrap these materials in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization, for example, hydroxymethylcellulose or gelatin-microcapsules and poly(methylmethacrylate) microcapsules, respectively, or in colloidal drug delivery systems, for example, liposomes, albumin microspheres, microemulsions, nanoparticles, and nanocapsules or in macroemulsions.

[00126] When oral preparations are desired, the compositions may be combined with typical carriers, such as lactose, sucrose, starch, talc magnesium stearate, crystalline cellulose, methyl cellulose, carboxymethyl cellulose, glycerin, sodium alginate or gum arabic among others.

7. Diagnosis, Treatment and Screening Methods using E-cadherin Nucleic

Acids

a. General Aspects

[00127] The nucleic acid molecules of the present invention are useful for probes, primers, chemical intermediates, and in biological assays. The nucleic acid molecules are useful as a hybridization probe for messenger RNA, transcript/cDNA and genomic DNA to detect or isolate full-length cDNA and genomic clones encoding E-cadherin protein or peptide of the invention, or variants thereof

[00128] The probe can correspond to any sequence along the entire length of the nucleic acid molecules of SEQ ID NOs: 4, 5 or 6. Accordingly, it could be derived from 5' noncoding regions, the coding region, and 3' noncoding regions.

[00129] The nucleic acid molecules are also useful as primers for PCR to amplify any given region of a nucleic acid molecule and are useful to synthesize antisense molecules of desired length and sequence.

[00130] The nucleic acid molecules are also useful for constructing recombinant vectors. Such vectors include expression vectors that express a portion of, or all of, the peptide sequences. The nucleic acid molecules are also useful for expressing antigenic portions of the proteins.

[00131] The nucleic acid molecules are also useful for designing ribozymes corresponding to all, or a part, of the mRNA produced from the nucleic acid molecules described herein.

[00132] The nucleic acid molecules are also useful for constructing host cells expressing a part, or all, of the nucleic acid molecules and peptides.

[00133] The nucleic acid molecules are also useful for constructing transgenic animals expressing all, or a part, of the nucleic acid molecules and peptides.

[00134] In vitro techniques for detection of mRNA include Northern hybridizations and in situ hybridizations. In vitro techniques for detecting DNA include Southern hybridizations and in situ hybridization.

b. Diagnosis methods

[00135] The nucleic acid molecules are also useful as hybridization probes for determining the presence, level, form and distribution of nucleic acid expression. The probes can be used to detect the presence of, or to determine levels of, a specific nucleic acid molecule in cells, tissues, and in organisms. Accordingly, probes corresponding to the peptides described herein can be used to assess expression and/or gene copy number in a given cell, tissue, or organism. These uses are relevant for diagnosis of disorders involving an increase or decrease in E-cadherin protein expression relative to normal results.

[00136] Probes can be used as a part of a diagnostic test kit for identifying cells or tissues that express E-cadherin protein differentially, such as by measuring a level of a E-cadherin-encoding nucleic acid in a sample of cells from a subject e.g., mRNA or genomic DNA, or determining if a E-cadherin gene has been mutated.

[00137] The invention also encompasses kits for detecting the presence of E-cadherin nucleic acid in a biological sample. For example, the kit can comprise reagents such as a labeled or labelable nucleic acid or agent capable of detecting E-cadherin nucleic acid in a biological sample; means for determining the amount of E-cadherin nucleic acid in the sample; and means for comparing the amount of E-cadherin nucleic acid in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect E-cadherin protein mRNA or DNA.

c. *Screening Method Using Nucleic Acids*

[00138] Nucleic acid expression assays are useful for drug screening to identify compounds that modulate E-cadherin nucleic acid expression.

[00139] The invention thus provides a method for identifying a compound that can be used to treat a pancreatic tumor or cancer associated with differential expression of the E-cadherin gene. The method typically includes assaying the ability of the compound to modulate the expression of E-cadherin nucleic acid and thus identifying a compound that can be used to treat a disorder characterized by undesired E-cadherin nucleic acid

expression. The assays can be performed in cell-based and cell-free systems. Cell-based assays include cells naturally expressing E-cadherin nucleic acid or recombinant cells genetically engineered to express specific nucleic acid sequences.

[00140] The assay for E-cadherin nucleic acid expression can involve direct assay of nucleic acid levels, such as mRNA levels, or on collateral compounds involved in the signal pathway. Further, the expression of genes that are up- or down-regulated in response to the E-cadherin protein signal pathway can also be assayed. In this embodiment the regulatory regions of these genes can be operably linked to a reporter gene such as luciferase.

[00141] Thus, modulators of E-cadherin gene expression can be identified in a method wherein a cell is contacted with a candidate compound or agent and the expression of mRNA determined. The level of expression of E-cadherin mRNA in the presence of the candidate compound or agent is compared to the level of expression of E-cadherin mRNA in the absence of the candidate compound or agent. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. When expression of mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of nucleic acid expression. When nucleic acid expression is statistically significantly less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of nucleic acid expression.

d. Methods of Monitoring Treatment

[00142] The nucleic acid molecules are also useful for monitoring the effectiveness of modulating compounds or agents on the expression or activity of the E-cadherin gene in clinical trials or in a treatment regimen. Thus, the gene expression pattern can serve as a barometer for the continuing effectiveness of treatment with the compound, particularly with compounds to which a patient can develop resistance. The gene expression pattern can also serve as a marker indicative of a physiological response of the affected cells to the compound. Accordingly, such monitoring would allow either increased

administration of the compound or the administration of alternative compounds to which the patient has not become resistant. Similarly, if the level of nucleic acid expression falls below a desirable level, administration of the compound could be commensurately decreased.

e. Treatment Using Nucleic Acid

[00143] The nucleic acid molecules are useful to design antisense constructs to control E-cadherin gene expression in cells, tissues, and organisms. A DNA antisense nucleic acid molecule is designed to be complementary to a region of the gene involved in transcription, preventing transcription and hence production of E-cadherin protein. An antisense RNA or DNA nucleic acid molecule would hybridize to the mRNA and thus block translation of mRNA into E-cadherin protein.

[00144] The nucleic acid of the present invention may also be used to specifically suppress gene expression by methods such as RNA interference (RNAi), which may also include cosuppression and quelling. This and antisense RNA or DNA of gene suppression are well known in the art. A review of this technique is found in Science 288:1370-1372, 2000. RNAi also operates on a post-transcriptional level and is sequence specific, but suppresses gene expression far more efficiently than antisense RNA. RNAi fragments, particularly double-stranded (ds) RNAi, can be also used to generate loss-of-function phenotypes.

[00145] Alternatively, a class of antisense molecules can be used to inactivate mRNA in order to decrease expression of E-cadherin nucleic acid. Accordingly, these molecules can treat a disorder characterized by abnormal or undesired E-cadherin nucleic acid expression. This technique involves cleavage by means of ribozymes containing nucleotide sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA to be translated. Possible regions include coding regions and particularly coding regions corresponding to the catalytic and other functional activities of the E-cadherin protein, such as substrate binding.

[00146] The nucleic acid molecules can be used for gene therapy in patients containing cells that are aberrant in E-cadherin gene expression. Thus, recombinant cells, which include the patient's cells that have been engineered ex vivo and returned to the patient, are introduced into an individual where the cells produce the desired E-cadherin protein to treat the individual.

8. Diagnosis using E-cadherin Protein

Protein detections

[00147] The present invention provides methods for diagnosing or detecting the differential presence of E-cadherin protein. Where E-cadherin is overexpressed in diseased cells, E-cadherin protein is detected directly.

[00148] The information obtained is also used to determine prognosis and appropriate course of treatment. For example, it is contemplated that individuals with a specific E-cadherin expression or stage of pancreatic diseases may respond differently to a given treatment than individuals lacking E-cadherin expression. The information obtained from the diagnostic methods of the present invention thus provides for the personalization of diagnosis and treatment.

[00149] In one embodiment, the present invention provides a method for monitoring pancreatic diseases treatment in a subject comprising: determining the level of E-cadherin protein or any fragment(s) or peptide(s) thereof in a test sample from said subject, wherein a level of said E-cadherin protein similar to the level of said protein in a test sample from a healthy subject, or the level established for a healthy subject, is indicative of successful treatment.

[00150] In another embodiment, the present invention provides a method for diagnosing recurrence of pancreatic diseases following successful treatment in a subject comprising: determining the level of E-cadherin protein or any fragment(s) or peptide(s) thereof in a test sample from said subject; wherein a changed level of said E-cadherin protein relative to the level of said protein in a test sample from a healthy

subject, or the level established for a healthy subject, is indicative of recurrence of pancreatic diseases.

[00151] In yet another embodiment, the present invention provides a method for diagnosing or detecting pancreatic diseases in a subject comprising: determining the level of E-cadherin protein or any fragment or peptides thereof in a test sample from said subject; wherein a differential level of said E-cadherin protein relative to the level of said protein in a test sample from a healthy subject, or the level established for a healthy subject, is indicative of pancreatic diseases.

[00152] These methods are also useful for diagnosing diseases that show differential protein expression. As describe earlier, normal, control or standard values or level established from a healthy subject for protein expression are established by combining body fluids or tissue, cell extracts taken from a normal healthy mammalian or human subject with specific antibodies to a protein under conditions for complex formation. Standard values for complex formation in normal and diseased tissues are established by various methods, often photometric means. Then complex formation as it is expressed in a subject sample is compared with the standard values. Deviation from the normal standard and toward the diseased standard provides parameters for disease diagnosis or prognosis while deviation away from the diseased and toward the normal standard may be used to evaluate treatment efficacy.

[00153] In yet another embodiment, the present invention provides a detection or diagnostic method of E-cadherin by using LC/MS. The proteins from cells are prepared by methods known in the art (for example, R. Aebersold, Nature Biotechnology, Volume 21, Number 6, June 2003). The differential expression of proteins in disease and healthy samples are quantitated using Mass Spectrometry and ICAT (Isotope Coded Affinity Tag) labeling, which is known in the art. ICAT is an isotope label technique that allows for discrimination between two populations of proteins, such as a healthy and a disease sample. The LC/MS spectra are collected for the labeled samples. The raw scans from the LC/MS instrument are subjected to peak detection and noise reduction software. Filtered peak lists are then used to detect 'features' corresponding to specific peptides from the

original sample(s). Features are characterized by their mass/charge, charge, retention time, isotope pattern and intensity.

[00154] The intensity of a peptide present in both healthy and disease samples can be used to calculate the differential expression, or relative abundance, of the peptide. The intensity of a peptide found exclusively in one sample can be used to calculate a theoretical expression ratio for that peptide (singleton). Expression ratios are calculated for each peptide of each replicate of the experiment. Thus overexpression or under expression of E-cadherin protein or peptide are similar to the expression pattern in a test subject indicates the likelihood of having pancreatic diseases or diseases associated with pancreas.

[00155] Immunological methods for detecting and measuring complex formation as a measure of protein expression using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), fluorescence-activated cell sorting (FACS) and antibody arrays. Such immunoassays typically involve the measurement of complex formation between the protein and its specific antibody. These assays and their quantitation against purified, labeled standards are well known in the art (Ausubel, supra, unit 10.1-10.6). A two-site, monoclonal-based immunoassay utilizing antibodies reactive to two non-interfering epitopes is preferred, but a competitive binding assay may be employed (Pound (1998) Immunochemical Protocols, Humana Press, Totowa N.J.). More immunological detections are described in section below.

[00156] For diagnostic applications, the antibody or its variant typically will be labeled with a detectable moiety. Numerous labels are available which can be generally grouped into the following categories:

(a) Radioisotopes, such as ^{36}S , ^{14}C , ^{125}I , ^3H , and ^{131}I . The antibody variant can be labeled with the radioisotope using the techniques described in Current Protocols in Immunology, vol 1-2, Coligen et al., Ed., Wiley-Interscience, New York, Pubs. (1991) for example and radioactivity can be measured using scintillation counting.

(b) Fluorescent labels such as rare earth chelates (europium chelates) or fluorescein and its derivatives, rhodamine and its derivatives, dansyl, Lissamine, phycoerythrin and Texas Red are available. The fluorescent labels can be conjugated to the antibody variant using the techniques disclosed in Current Protocols in Immunology, supra, for example. Fluorescence can be quantified using a fluorometer.

(c) Various enzyme-substrate labels are available and U.S. Pat. Nos. 4,275,149 and 4,318,980 provide a review of some of these. The enzyme generally catalyzes a chemical alteration of the chromogenic substrate that can be measured using various techniques. For example, the enzyme may catalyze a color change in a substrate, which can be measured spectrophotometrically. Alternatively, the enzyme may alter the fluorescence or chemiluminescence of the substrate. Techniques for quantifying a change in fluorescence are described above. The chemiluminescent substrate becomes electronically excited by a chemical reaction and may then emit light which can be measured (using a chemiluminometer, for example) or donates energy to a fluorescent acceptor. Examples of enzymatic labels include luciferases (e.g., firefly luciferase and bacterial luciferase; U.S. Pat. No. 4,737,456), luciferin, 2,3-dihydrophthalazinediones, malate dehydrogenase, urease, peroxidase such as horseradish peroxidase (HRPO), alkaline phosphatase, β -galactosidase, glucoamylase, lysozyme, saccharide oxidases (e.g., glucose oxidase, galactose oxidase, and glucose-6-phosphate dehydrogenase), heterocyclic oxidases (such as uricase and xanthine oxidase), lactoperoxidase, microperoxidase, and the like. Techniques for conjugating enzymes to antibodies are described in O'Sullivan et al., Methods for the Preparation of Enzyme-Antibody Conjugates for Use in Enzyme Immunoassay, in Methods in Enzyme. (Ed. J. Langone & H. Van Vunakis), Academic press, New York, 73: 147-166 (1981).

[00157] Sometimes, the label is indirectly conjugated with the antibody. The skilled artisan will be aware of various techniques for achieving this. For example, the antibody can be conjugated with biotin and any of the three broad categories of labels mentioned above can be conjugated with avidin, or vice versa. Biotin binds selectively to avidin and thus, the label can be conjugated with the antibody in this indirect manner. Alternatively, to achieve indirect conjugation of the label with the antibody, the antibody is conjugated

with a small hapten (e.g. digoxin) and one of the different types of labels mentioned above is conjugated with an anti-hapten antibody (e.g. anti-digoxin antibody). Thus, indirect conjugation of the label with the antibody can be achieved.

[00158] The biological samples can then be tested directly for the presence of E-cadherin by assays (e.g., ELISA or radioimmunoassay) and format (e.g., microwells, dipstick, etc., as described in International Patent Publication WO 93/03367). Alternatively, proteins in the sample can be size separated (e.g., by polyacrylamide gel electrophoresis (PAGE)), in the presence or absence of sodium dodecyl sulfate (SDS), and the presence of E-cadherin detected by immunoblotting (e.g., Western blotting). Immunoblotting techniques are generally more effective with antibodies generated against a peptide corresponding to an epitope of a protein, and hence, are particularly suited to the present invention.

[00159] Antibody binding may be detected also by "sandwich" immunoassays, immunoradiometric assays, gel diffusion precipitation reactions, immunodiffusion assays, in situ immunoassays (e.g., using colloidal gold, enzyme or radioisotope labels, for example), precipitation reactions, agglutination assays (e.g., gel agglutination assays, hemagglutination assays, etc.), complement fixation assays, immunofluorescence assays, protein A assays, and immunoelectrophoresis assays, etc.

[00160] In one embodiment, antibody binding is detected by detecting a label on the primary antibody. In another embodiment, the primary antibody is detected by detecting binding of a secondary antibody or reagent to the primary antibody. In a further embodiment, the secondary antibody is labeled. Many means are known in the art for detecting binding in an immunoassay and are within the scope of the present invention. As is well known in the art, the immunogenic peptide should be provided free of the carrier molecule used in any immunization protocol. For example, if the peptide is conjugated to KLH, it may be conjugated to BSA, or used directly, in a screening assay. In some embodiments, an automated detection assay is utilized. Methods for the automation of immunoassays are well known in the art (See e.g., U.S. Pat. Nos. 5,885,530, 4,981,785, 6,159,750, and 5,358,691, each of which is herein incorporated by

reference). In some embodiments, the analysis and presentation of results is also automated. For example, in some embodiments, software that generates a prognosis based on the presence or absence of a series of antigens is utilized.

[00161] Competitive binding assays rely on the ability of a labeled standard to compete with the test sample for binding with a limited amount of antibody. The amount of antigen in the test sample is inversely proportional to the amount of standard that becomes bound to the antibodies. To facilitate determining the amount of standard that becomes bound, the antibodies generally are insolubilized before or after the competition. As a result, the standard and test sample that are bound to the antibodies may conveniently be separated from the standard and test sample, which remain unbound.

[00162] Sandwich assays involve the use of two antibodies, each capable of binding to a different immunogenic portion, or epitope, or the protein to be detected. In a sandwich assay, the test sample to be analyzed is bound by a first antibody, which is immobilized on a solid support, and thereafter a second antibody binds to the test sample, thus forming an insoluble three-part complex. See e.g., U.S. Pat. No. 4,376,110. The second antibody may itself be labeled with a detectable moiety (direct sandwich assays) or may be measured using an anti-immunoglobulin antibody that is labeled with a detectable moiety (indirect sandwich assay). For example, one type of sandwich assay is an ELISA assay, in which case the detectable moiety is an enzyme.

[00163] The antibodies may also be used for in vivo diagnostic assays. Generally, the antibody is labeled with a radionucleotide (such as ^{111}In , ^{99}Tc , ^{14}C , ^{131}I , ^3H , ^{32}P or ^{35}S) so that the tumor can be localized using immunoscintigraphy. In one embodiment, antibodies or fragments thereof bind to the extracellular domains of two or more E-cadherin targets and the affinity value(Kd) is less than $1 \times 10^8 \text{ M}$.

[00164] Antibodies for diagnostic use may be labeled with probes suitable for detection by various imaging methods. Methods for detection of probes include, but are not limited to, fluorescence, light, confocal and electron microscopy; magnetic resonance imaging and spectroscopy; fluoroscopy, computed tomography and positron emission tomography. Suitable probes include, but are not limited to, fluorescein, rhodamine,

eosin and other fluorophores, radioisotopes, gold, gadolinium and other lanthanides, paramagnetic iron, fluorine-18 and other positron-emitting radionuclides. Additionally, probes may be bi- or multi-functional and be detectable by more than one of the methods listed. These antibodies may be directly or indirectly labeled with said probes. Attachment of probes to the antibodies includes covalent attachment of the probe, incorporation of the probe into the antibody, and the covalent attachment of a chelating compound for binding of probe, amongst others well recognized in the art.

[00165] For immunohistochemistry, the disease tissue sample may be fresh or frozen or may be embedded in paraffin and fixed with a preservative such as formalin (see Example). The fixed or embedded section contains the sample are contacted with a labeled primary antibody and secondary antibody, wherein the antibody is used to detect E-cadherin protein expression *in situ*. The detailed procedure is shown in the Example.

[00166] Antibodies against E-cadherin protein or peptides are useful to detect the presence of one of the proteins of the present invention in cells or tissues to determine the pattern of expression of the protein among various tissues in an organism and over the course of normal development.

[00167] Further, such antibodies can be used to detect protein *in situ*, *in vitro*, or in a cell lysate or supernatant in order to evaluate the abundance and pattern of expression. Also, such antibodies can be used to assess abnormal tissue distribution or abnormal expression during development or progression of a biological condition. Antibody detection of circulating fragments of the full length protein can be used to identify turnover.

[00168] Further, the antibodies can be used to assess expression in disease states such as in active stages of the disease or in an individual with a predisposition toward disease related to the protein's function. When a disorder is caused by an inappropriate tissue distribution, developmental expression, level of expression of the protein, or expressed/processed form, the antibody can be prepared against the normal protein. Experimental data as provided in Table 1 indicates expression in human pancreatic cell lines. If a disorder is characterized by a specific mutation in the protein, antibodies

specific for this mutant protein can be used to assay for the presence of the specific mutant protein.

[00169] The antibodies can also be used to assess normal and aberrant subcellular localization of cells in the various tissues in an organism. Experimental data as provided in Table 1 indicates expression in human pancreatic cell lines. The diagnostic uses can be applied, not only in genetic testing, but also in monitoring a treatment modality. Accordingly, where treatment is ultimately aimed at correcting expression level or the presence of aberrant sequence and aberrant tissue distribution or developmental expression, antibodies directed against the protein or relevant fragments can be used to monitor therapeutic efficacy. More detection and diagnostic methods are described in detail below.

[00170] Additionally, antibodies are useful in pharmacogenomic analysis. Thus, antibodies prepared against polymorphic proteins can be used to identify individuals that require modified treatment modalities. The antibodies are also useful as diagnostic tools, as an immunological marker for aberrant protein analyzed by electrophoretic mobility, isoelectric point, tryptic peptide digest, and other physical assays known to those in the art.

[00171] The antibodies are also useful for tissue typing. Where a specific protein has been correlated with expression in a specific tissue, antibodies that are specific for this protein can be used to identify a tissue type.

[00172] The invention also encompasses kits for using antibodies to detect the presence of a protein in a biological sample. The kit can comprise antibodies such as a labeled or labelable antibody and a compound or agent for detecting protein in a biological sample; means for determining the amount of protein in the sample; means for comparing the amount of protein in the sample with a standard; and instructions for use. Such a kit can be supplied to detect a single protein or epitope or can be configured to detect one of a multitude of epitopes, such as in an antibody detection array. Arrays are described in detail below for nucleic acid arrays and similar methods have been developed for antibody arrays.

9. Methods of Treatment Based on E-cadherin Protein

a. Antibody therapy

[00173] The antibody of the present invention can be used for therapeutic reasons. It is contemplated that the antibody of the present invention may be used to treat a mammal, preferably a human with a pancreatic disease.

[00174] In general, the antibodies are also useful for inhibiting protein function, for example, blocking the binding of E-cadherin protein or peptide to a binding partner such as a substrate. These uses can also be applied in a therapeutic context in which treatment involves inhibiting the protein's function. An antibody can be used, for example, to block binding, thus modulating (agonizing or antagonizing) the peptides activity. Antibodies can be prepared against specific fragments containing sites required for function or against intact protein that is associated within a cell or cell membrane. The functional blocking assays are provided in detail in the Examples.

[00175] The antibodies of present invention can also be used as means of enhancing the immune response. The antibodies can be administered in amounts similar to those used for other therapeutic administrations of antibody. For example, pooled gamma globulin is administered at a range of about 1 mg to about 100 mg per patient.

[00176] Antibodies reactive with the protein or peptides of E-cadherin can be administered alone or in conjunction with other anti-cancer therapies to a mammal afflicted-with pancreatic diseases or cancer. Examples of anti-cancer therapies include, but are not limited to, chemotherapy, radiation therapy, and adoptive immunotherapy therapy with TIL (Tumor Infiltration Lymphocytes).

[00177] The selection of an antibody subclass for therapy will depend upon the nature of the antigen to be acted upon. For example, an IgM may be preferred in situations where the antigen is highly specific for the diseased target and rarely occurs on normal cells. However, where the disease-associated antigen is also expressed in normal tissues, although at much lower levels, the IgG subclass may be preferred, since the binding of at least two IgG molecules in close proximity is required to activate complement, less

complement mediated damage may occur in the normal tissues which express smaller amounts of the antigen and, therefore, bind fewer IgG antibody molecules. Furthermore, IgG molecules by being smaller may be more able than IgM molecules to localize to the diseased tissue.

[00178] The mechanism for antibody therapy is that the therapeutic antibody recognizes a cell surface protein or a cytosolic protein that is expressed or preferably, overexpressed in a diseased cell. By NK cell or complement activation, or conjugation of the antibody with an immunotoxin or radiolabel, the interaction can abrogate ligand/receptor interaction or activation of apoptosis.

[00179] The potential mechanisms of antibody-mediated cytotoxicity of diseased cells are phagocyte (antibody dependent cellular cytotoxicity (ADCC)) (see Example), complement (Complement-mediated cytotoxicity (CMC)) (see Example), naked antibody (receptor cross-linking apoptosis and growth factor inhibition), or targeted payload labeled with radionuclide or immunotoxins or immunochemotherapeutics.

[00180] In one embodiment, the antibody is administered to a nonhuman mammal for the purposes of obtaining preclinical data, for example. Exemplary nonhuman mammals to be treated include nonhuman primates, dogs, cats, rodents and other mammals in which preclinical studies are performed. Such mammals may be established animal models for a disease to be treated with the antibody or may be used to study toxicity of the antibody of interest. In each of these embodiments, dose escalation studies may be performed on the mammal.

[00181] The antibody is administered by any suitable means, including parenteral, subcutaneous, intraperitoneal, intrapulmonary, and intranasal, and, if desired for local immunosuppressive treatment, intralesional administration. Parenteral infusions include intramuscular, intravenous, intraarterial, intraperitoneal, or subcutaneous administration. In addition, the antibody variant is suitably administered by pulse infusion, particularly with declining doses of the antibody variant. Preferably the dosing is given by injections, most preferably intravenous or subcutaneous injections, depending in part on whether the administration is brief or chronic.

[00182] For the prevention or treatment of a disease, the appropriate dosage of the antibody will depend on the type of disease to be treated, the severity and the course of the disease, whether the antibody is administered for preventive or therapeutic purposes, previous therapy, the patient's clinical history and response to the antibody, and the discretion of the attending physician.

[00183] Depending on the type and severity of the disease, about 1 µg/kg to 150 mg/kg (e.g., 0.1-20 mg/kg) of antibody is an initial candidate dosage for administration to the patient, whether, for example, by one or more separate administrations, or by continuous infusion. A typical daily dosage might range from about 1 µg/kg to 100 mg/kg or more, depending on the factors mentioned above. For repeated administrations over several days or longer, depending on the condition, the treatment is sustained until a desired suppression of disease symptoms occurs. However, other dosage regimens may be useful. The progress of this therapy is easily monitored by conventional techniques and assays.

[00184] The antibody composition will be formulated, dosed and administered in a manner consistent with good medical practice. Factors for consideration in this context include the particular disorder being treated, the particular mammal being treated, the clinical condition of the individual patient, the cause of the disorder, the site of delivery of the agent, the method of administration, the scheduling of administration, and other factors known to medical practitioners.

[00185] The therapeutically effective amount of the antibody to be administered will be governed by such considerations, and is the minimum amount necessary to prevent, ameliorate, or treat a disease or disorder. The antibody may optionally be formulated with one or more agents currently used to prevent or treat the disorder in question.

[00186] Suitable agents in this regard include radionuclides, differentiation inducers, drugs, toxins, and derivatives thereof. Preferred radionuclides include ^{90}Y , ^{123}I , ^{125}I , ^{131}I , ^{186}Re , ^{188}Re , ^{211}At , and ^{212}Bi . Preferred drugs include methotrexate, and pyrimidine and purine analogs. Preferred differentiation inducers include phorbol esters and butyric acid.

Preferred toxins include ricin, abrin, diphtheria toxin, cholera toxin, gelonin, *Pseudomonas* exotoxin, *Shigella* toxin, and pokeweed antiviral protein

[00187] A therapeutic agent may be coupled (e.g., covalently bonded) to a suitable antibody either directly or indirectly (e.g., via a linker group). A direct reaction between an agent and an antibody is possible when each possesses a substituent capable of reacting with the other. For example, a nucleophilic group, such as an amino or sulfhydryl group, on one may be capable of reacting with a carbonyl-containing group, such as an anhydride or an acid halide, or with an alkyl group containing a good leaving group (e.g., a halide) on the other.

[00188] Alternatively, it may be desirable to couple a therapeutic agent and an antibody via a linker group. A linker group can function as a spacer to distance an antibody from an agent in order to avoid interference with binding capabilities. A linker group can also serve to increase the chemical reactivity of a substituent on an agent or an antibody, and thus increase the coupling efficiency. An increase in chemical reactivity may also facilitate the use of agents, or functional groups on agents, which otherwise would not be possible.

[00189] It will be evident to those skilled in the art that a variety of bifunctional or polyfunctional reagents, both homo- and hetero-functional (such as those described in the catalog of the Pierce Chemical Co., Rockford, Ill.), may be employed as the linker group. Coupling may be affected, for example, through amino groups, carboxyl groups, sulfhydryl groups or oxidized carbohydrate residues. There are numerous references describing such methodology, e.g. U.S. Pat. No. 4,671,958, to Rodwell et al.

[00190] Where a therapeutic agent is more potent when free from the antibody portion of the immunoconjugates of the present invention, it may be desirable to use a linker group which is cleavable during or upon internalization into a cell. A number of different cleavable linker groups have been described. The mechanisms for the intracellular release of an agent from these linker groups include cleavage by reduction of a disulfide bond (e.g., U.S. Pat. No. 4,489,710, to Spitler), by irradiation of a photolabile bond (e.g., U.S. Pat. No. 4,625,014, to Senter et al.), by hydrolysis of derivatized amino acid side chains

(e.g., U.S. Pat. No. 4,638,045, to Kohn et al.), by serum complement-mediated hydrolysis (e.g., U.S. Pat. No. 4,671,958, to Rodwell et al.), and acid-catalyzed hydrolysis (e.g., U.S. Pat. No. 4,569,789, to Blattler et al.).

[00191] It may be desirable to couple more than one agent to an antibody. In one embodiment, multiple molecules of an agent are coupled to one antibody molecule. In another embodiment, more than one type of agent may be coupled to one antibody. Regardless of the particular embodiment, immunoconjugates with more than one agent may be prepared in a variety of ways as described above.

b. *Other immunotherapy*

[00192] Peptides derived from the E-cadherin protein sequence may be modified to increase their immunogenicity by enhancing the binding of the peptide to the MHC molecules in which the peptide is presented. The peptide or modified peptide may be conjugated to a carrier molecule to enhance the antigenicity of the peptide. Examples of carrier molecules, include, but are not limited to, human albumin, bovine albumin, lipoprotein and keyhole limpet hemo-cyanin ("Basic and Clinical Immunology" (1991) Stites, D. P. and Terr A. I. (eds) Appleton and Lange, Norwalk Conn., San Mateo, Calif.).

[00193] An "immunogenic peptide" is a peptide, which comprises an allele-specific motif such that the peptide will bind the MHC allele (HLA in human) and be capable of inducing a CTL (cytotoxic T-lymphocytes) response. Thus, immunogenic peptides are capable of binding to an appropriate class I or II MHC molecule and inducing a cytotoxic T cell or T helper cell response against the antigen from which the immunogenic peptide is derived.

[00194] Alternatively, amino acid sequence variants of the peptide can be prepared by altering the nucleic acid sequence of the DNA which encodes the peptide, or by peptide synthesis. At the genetic level, these variants ordinarily are prepared by site-directed mutagenesis of nucleotides in the DNA encoding the peptide molecule, thereby producing DNA encoding the variant, and thereafter expressing the DNA in recombinant

cell culture. The variants typically exhibit the same qualitative biological activity as the nonvariant peptide.

[00195] The recombinant or natural protein, peptides, or fragment thereof of E-cadherin, or modified peptides, may be used as a vaccine either prophylactically or therapeutically. When provided prophylactically the vaccine is provided in advance of any evidence of pancreatic diseases, particularly, cancer. The prophylactic administration of the pancreatic disease vaccine should serve to prevent or attenuate pancreatic diseases, preferably cancer, in a mammal.

[00196] Preparation of vaccine uses recombinant protein or peptide expression vectors comprising a nucleic acid sequence encoding all or part of the E-cadherin protein. Examples of vectors that may be used in the aforementioned vaccines include, but are not limited to, defective retroviral vectors, adenoviral vectors, vaccinia viral vectors, fowl pox viral vectors, or other viral vectors (Mulligan, R. C., (1993) Science 260:926-932). The vectors can be introduced into a mammal either prior to any evidence of the pancreatic diseases or to mediate regression of the disease in a mammal afflicted with pancreatic diseases. Examples of methods for administering the viral vector into the mammals include, but are not limited to, exposure of cells to the virus *ex vivo*, or injection of the retrovirus or a producer cell line of the virus into the affected tissue or intravenous administration of the virus. Alternatively the vector may be administered locally by direct injection into the cancer lesion or topical application in a pharmaceutically acceptable carrier. The quantity of viral vector, carrying all or part of the E-cadherin nucleic acid sequence, to be administered is based on the titer of virus particles. A preferred range may be about 10^6 to about 10^{11} virus particles per mammal, preferably a human.

[00197] After immunization the efficacy of the vaccine can be assessed by the production of antibodies or immune cells that recognize the antigen, as assessed by specific lytic activity or specific cytokine production or by tumor regression. One skilled in the art would know the conventional methods to assess the aforementioned parameters. If the mammal to be immunized is already afflicted with cancer, the vaccine can be administered in conjunction with other therapeutic treatments. Examples of other

therapeutic treatments includes, but are not limited to, adoptive T cell immunotherapy, coadministration of cytokines or other therapeutic drugs for cancer.

[00198] Alternatively all or parts thereof of a substantially or partially purified the E-cadherin protein or their peptides may be administered as a vaccine in a pharmaceutically acceptable carrier. Ranges of the protein that may be administered are about 0.001 to about 100 mg per patient, preferred doses are about 0.01 to about 100 mg per patient. Immunization may be repeated as necessary, until a sufficient titer of anti-immunogen antibody or immune cells has been obtained.

[00199] In yet another alternative embodiment a viral vector, such as a retroviral vector, can be introduced into mammalian cells. Examples of mammalian cells into which the retroviral vector can be introduced include, but are not limited to, primary mammalian cultures or continuous mammalian cultures, COS cells, NIH3T3, or 293 cells (ATTC #CRL 1573), dendritic cells. The means by which the vector carrying the gene may be introduced into a cell includes, but is not limited to, microinjection, electroporation, transfection or transfection using DEAE dextran, lipofection, calcium phosphate or other procedures known to one skilled in the art (Sambrook et al. 3rd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (2001).

[00200] The vaccine formulation of the present invention comprises an immunogen that induces an immune response directed against the cancer associated antigen such as E-cadherin protein, and in nonhuman primates and finally in humans. The safety of the immunization procedures is determined by looking for the effect of immunization on the general health of the immunized animal (weight change, fever, appetite behavior etc.) and looking for pathological changes on autopsies. After initial testing in animals, cancer patients can be tested. Conventional methods would be used to evaluate the immune response of the patient to determine the efficiency of the vaccine.

[00201] In one embodiment mammals, preferably human, at high risk for pancreatic diseases, particularly cancer, are prophylactically treated with the vaccines of this invention. Examples include, but are not limited to, humans with a family history of pancreatic diseases, humans with a history of pancreatic diseases, particular cancer, or

humans afflicted with pancreatic cancer previously resected and therefore at risk for reoccurrence. When provided therapeutically, the vaccine is provided to enhance the patient's own immune response to the diseased antigen present on the pancreatic diseases or advanced stage of pancreatic diseases. The vaccine, which acts as an immunogen, may be a cell, cell lysate from cells transfected with a recombinant expression vector, or a culture supernatant containing the expressed protein. Alternatively, the immunogen is a partially or substantially purified recombinant protein, peptide or analog thereof or modified peptides or analogs thereof. The proteins or peptides may be conjugated with lipoprotein or administered in liposomal form or with adjuvant.

[00202] While it is possible for the immunogen to be administered in a pure or substantially pure form, it is preferable to present it as a pharmaceutical composition, formulation or preparation, as discussed hereinabove.

[00203] Vaccination can be conducted by conventional methods. For example, the immunogen can be used in a suitable diluent such as saline or water, or complete or incomplete adjuvants. Further, the immunogen may or may not be bound to a carrier to make the protein immunogenic. Examples of such carrier molecules include but are not limited to bovine serum albumin (BSA), keyhole limpet hemocyanin (KLH), tetanus toxoid, and the like. The immunogen also may be coupled with lipoproteins or administered in liposomal form or with adjuvants. The immunogen can be administered by any route-appropriate for antibody production such as intravenous, intraperitoneal, intramuscular, subcutaneous, and the like. The immunogen may be administered once or at periodic intervals until a significant titer of anti-E-cadherin immune cells or anti-E-cadherin antibody is produced. The presence of anti-E-cadherin immune cells may be assessed by measuring the frequency of precursor CTL (cytotoxic T-lymphocytes) against E-cadherin antigen prior to and after immunization by a CTL precursor analysis assay (Coulie, P. et al., (1992) International Journal Of Cancer 50:289-297). The antibody may be detected in the serum using the immunoassay described above.

[00204] The safety of the immunization procedures is determined by examining the effect of immunization on the general health of the immunized animal (fever, change in

weight, appetite, behavior etc.) and pathological changes on autopsies. After initial testing in animals, human patients can be tested. Conventional methods would be used to evaluate the immune response of the patient to determine the efficiency of the vaccine.

[00205] In yet another embodiment of this invention all, part, or parts of the E-cadherin protein or peptides or fragments thereof, or modified peptides, may be exposed to dendritic cells cultured *in vitro*. The cultured dendritic cells provide a means of producing T-cell dependent antigens comprised of dendritic cell modified antigen or dendritic cells pulsed with antigen, in which the antigen is processed and expressed on the antigen activated dendritic cell. The E-cadherin antigen activated dendritic cells or processed dendritic cell antigens may be used as immunogens for vaccines or for the treatment of pancreatic diseases, particularly pancreatic cancer. The dendritic cells should be exposed to the antigen for sufficient time to allow the antigens to be internalized and presented on the dendritic cells surface. The resulting dendritic cells or the dendritic-cell processed antigens can then be administered to an individual in need of therapy. Such methods are described in Steinman et al. (WO93/208185) and in Banchereau et al. (EPO Application 0563485A1).

[00206] In yet another aspect of this invention T-cells isolated from individuals can be exposed to E-cadherin protein, peptides or fragment thereof, or modified peptides *in vitro* and then administered to a patient in need of such treatment in a therapeutically effective amount. Examples of where T-lymphocytes can be isolated include but are not limited to, peripheral blood cells lymphocytes (PBL), lymph nodes, or tumor infiltrating lymphocytes (TIL). Such lymphocytes can be isolated from the individual to be treated or from a donor by methods known in the art and cultured *in vitro* (Kawakami, Y. et al. (1989) J. Immunol. 142: 2453-3461). Lymphocytes are cultured in media such as RPMI or RPMI 1640 or AIM V for 1-10 weeks. Viability is assessed by trypan blue dye exclusion assay. Examples of how these sensitized T-cells can be administered to the mammal include but are not limited to, intravenously, intraperitoneally or intralesionally. Parameters that may be assessed to determine the efficacy of these sensitized T-lymphocytes include, but are not limited to, production of immune cells in the mammal being treated or tumor regression. Conventional methods are used to assess these

parameters. Such treatment can be given in conjunction with cytokines or gene modified cells (Rosenberg, S. A. et al. (1992) Human Gene Therapy, 3: 75-90; Rosenberg, S. A. et al. (1992) Human Gene Therapy, 3: 57-73).

[00207] The present invention is further described by the following examples, which are provided solely to illustrate the invention by reference to specific embodiments. This exemplification, while illustrating certain aspects of the invention, does not offer the limitations or circumscribe the scope of the disclosed invention.

10. Screening Methods Using Proteins

[00208] The E-cadherin protein and polypeptide can be used to identify compounds or agents that modulate E-cadherin activity of the protein in its natural state or an altered form that causes a specific disease or pathology associated with E-cadherin. Both E-cadherin of the present invention and appropriate variants and fragments can be used in high-throughput screens to assay candidate compounds for the ability to bind to E-cadherin. These compounds can be further screened against functional E-cadherin to determine the effect of the compound on E-cadherin activity. Further, these compounds can be tested in animal or invertebrate systems to determine activity/effectiveness. Compounds can be identified that activate (agonist) or inactivate (antagonist) E-cadherin to a desired degree.

[00209] Both E-cadherin of the present invention and appropriate variants and fragments can be used in high-throughput screening to assay candidate compounds for the ability to bind to E-cadherin. These compounds can be further screened against functional E-cadherin to determine the effect of the compound on E-cadherin activity. Further, these compounds can be tested in animal or invertebrate systems to determine activity/effectiveness. Compounds can be identified that activate (agonist) or inactivate (antagonist) E-cadherin to a desired degree.

[00210] Further, the proteins of the present invention can be used to screen a compound or an agent for the ability to stimulate or inhibit interaction between E-cadherin protein and a molecule that normally interacts with E-cadherin protein, e.g. a

substrate or an extracellular binding ligand or a component of the signal pathway that E-cadherin protein normally interacts (for example, a cytosolic signal protein). Such assays typically include the steps of combining E-cadherin protein with a candidate compound under conditions that allow E-cadherin protein, or fragment, to interact with the target molecule, and to detect the formation of a complex between the protein and the target or to detect the biochemical consequence of the interaction with E-cadherin protein and the target, such as any of the associated effects of signal transduction such as protein phosphorylation, cAMP turnover, and adenylate cyclase activation, etc.

[00211] Candidate compounds or agents include 1) peptides such as soluble peptides, including Ig-tailed fusion peptides and members of random peptide libraries (see, e.g., Lam et al., *Nature* 354:82-84 (1991); Houghten et al., *Nature* 354:84-86 (1991)) and combinatorial chemistry-derived molecular libraries made of D- and/or L- configuration amino acids; 2) phosphopeptides (e.g., members of random and partially degenerate, directed phosphopeptide libraries, see, e.g., Songyang et al., *Cell* 72:767-778 (1993)); 3) antibodies (e.g., polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single chain antibodies as well as Fab, F(ab')₂, Fab expression library fragments, and epitope-binding fragments of antibodies); and 4) small organic and inorganic molecules (e.g., molecules obtained from combinatorial and natural product libraries).

[00212] One candidate compound or agent is a soluble fragment of E-cadherin that competes for substrate binding. Other candidate compounds include mutant E-cadherin or appropriate fragments containing mutations that affect E-cadherin function and thus compete for substrate. Accordingly, a fragment that competes for substrate, for example with a higher affinity, or a fragment that binds substrate but does not allow release, is encompassed by the invention.

[00213] Any of the biological or biochemical functions mediated by E-cadherin can be used as an endpoint assay to identify an agent that modulates E-cadherin activity. These include all of the biochemical or biochemical/biological events described herein, in the references cited herein, incorporated by reference for these endpoint assay targets, and other functions known to those of ordinary skill in the art or that can be readily identified.

Specifically, a biological function of a cell or tissues that expresses E-cadherin can be assayed.

[00214] A substrate-binding region can be used that interacts with a different substrate than one that is recognized by the native E-cadherin. Accordingly, a different set of signal transduction components is available as an end-point assay for activation. This allows for assays to be performed in other than the specific host cell from which E-cadherin is derived.

[00215] Competition binding assays may also be used to discover compounds that interact with E-cadherin (e.g. binding partners and/or ligands). Thus, a compound is exposed to E-cadherin polypeptide under conditions that allow the compound to bind or to otherwise interact with the polypeptide. Soluble E-cadherin polypeptide is also added to the mixture. If the test compound interacts with the soluble E-cadherin polypeptide, it decreases the amount of complex formed or activity from E-cadherin. This type of assay is particularly useful in cases in which compounds are sought that interact with specific regions of E-cadherin. Thus, the soluble polypeptide that competes with the target E-cadherin region is designed to contain peptide sequences corresponding to the region of interest.

[00216] To perform cell free drug screening assays, it is sometimes desirable to immobilize either the E-cadherin protein, or fragment, or its target molecule to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay.

[00217] Techniques for immobilizing proteins on matrices can be used in the drug screening assays. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase fusion proteins can be adsorbed onto glutathione SEPHAROSE beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the cell lysates (e.g., ³⁵S-labeled) and the candidate compound, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads are washed to

remove any unbound label, and the matrix immobilized and radiolabel determined directly, or in the supernatant after the complexes are dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of E-cadherin-binding protein found in the bead fraction quantitated from the gel using standard electrophoretic techniques. For example, either the polypeptide or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin using techniques well known in the art. Alternatively, antibodies reactive with the protein but which do not interfere with binding of the protein to its target molecule can be derivatized to the wells of the plate, and the protein trapped in the wells by antibody conjugation. Preparations of E-cadherin-binding protein and a candidate compound are incubated in E-cadherin protein-presenting wells and the amount of complex trapped in the well can be quantitated. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the E-cadherin protein target molecule, or which are reactive with E-cadherin protein and compete with the target molecule, as well as E-cadherin-linked assays which rely on detecting an enzymatic activity associated with the target molecule.

[00218] Agents that modulate E-cadherin of the present invention can be identified using one or more of the above assays, alone or in combination. It is generally preferable to use a cell-based or cell free system first and then confirm activity in an animal or other model system. Such model systems are well known in the art and can readily be employed in this context.

[00219] In yet another aspect of the invention, E-cadherin protein can be used as a "bait protein" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos et al. (1993) *Cell* 72:223-232; Madura et al. (1993) *J. Biol. Chem.* 268:12046-12054; Bartel et al. (1993) *Biotechniques* 14:920-924; Iwabuchi et al. (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with E-cadherin and are involved in E-cadherin activity. Such E-cadherin-binding proteins are also likely to be involved in the propagation of signals by E-cadherin protein or E-cadherin targets as, for example, downstream elements of a E-cadherin-

mediated signaling pathway. Alternatively, such E-cadherin -binding proteins are likely to be E-cadherin inhibitors.

[00220] The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for E-cadherin protein is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences that encode an unidentified protein (“prey” or “sample”) is fused to a gene that codes for the activation domain of the known transcription factor. If the “bait” and the “prey” proteins are able to interact, in vivo, forming a E-cadherin-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with E-cadherin protein.

Array:

[00221] “Array” refers to an ordered arrangement of at least two transcripts, proteins or peptides, or antibodies on a substrate. At least one of the transcripts, proteins, or antibodies represents a control or standard, and the other transcript, protein, or antibody is of diagnostic or therapeutic interest. The arrangement of at least two and up to about 40,000 transcripts, proteins, or antibodies on the substrate assures that the size and signal intensity of each labeled complex, formed between each transcript and at least one nucleic acid, each protein and at least one ligand or antibody, or each antibody and at least one protein to which the antibody specifically binds, is individually distinguishable.

[00222] An “expression profile” is a representation of gene expression in a sample. A nucleic acid expression profile is produced using sequencing, hybridization, or amplification technologies using transcripts from a sample. A protein expression profile, although time delayed, mirrors the nucleic acid expression profile and is produced using

gel electrophoresis, mass spectrometry, or an array and labeling moieties or antibodies which specifically bind the protein. The nucleic acids, proteins, or antibodies specifically binding the protein may be used in solution or attached to a substrate, and their detection is based on methods well known in the art.

[00223] A substrate includes but is not limited to, paper, nylon or other type of membrane, filter, chip, glass slide, or any other suitable solid support.

[00224] The present invention also provides an antibody array. Antibody arrays have allowed the development of techniques for high-throughput screening of recombinant antibodies. Such methods use robots to pick and grid bacteria containing antibody genes, and a filter-based ELISA to screen and identify clones that express antibody fragments. Because liquid handling is eliminated and the clones are arrayed from master stocks, the same antibodies can be spotted multiple times and screened against multiple antigens simultaneously. For more information, see de Wildt et al. (2000) *Nat. Biotechnol.* 18:989-94.

[00225] The array is prepared and used according to the methods described in U.S. Patent No. 5,837,832, Chee et al., PCT application W095/11995 (Chee et al.), Lockhart, D. J. et al. (1996; *Nat. Biotech.* 14: 1675-1680) and Schena, M. et al. (1996; *Proc. Natl. Acad. Sci.* 93: 10614-10619), US Patent No. 5,807,522, Brown et al., all of which are incorporated herein in their entirety by reference.

[00226] In one embodiment, a nucleic acid array or a microarray, preferably composed of a large number of unique, single-stranded nucleic acid sequences, usually either synthetic antisense oligonucleotides or fragments of cDNAs, fixed to a solid support. The oligonucleotides are preferably about 6-60 nucleotides in length, more preferably 15-30 nucleotides in length, and most preferably about 20-25 nucleotides in length.

[00227] In order to produce oligonucleotides to a known sequence for an array, the gene(s) of interest (or an ORF identified from the contigs of the present invention) is typically examined using a computer algorithm which starts at the 5' or at the 3' end of the nucleotide sequence. Typical algorithms will then identify oligomers of defined

length that are unique to the gene, have a GC content within a range suitable for hybridization, and lack predicted secondary structure that may interfere with hybridization. In certain situations it may be appropriate to use pairs of oligonucleotides on an array. The "pairs" will be identical, except for one nucleotide that preferably is located in the center of the sequence. The second oligonucleotide in the pair (mismatched by one) serves as a control. The number of oligonucleotide pairs may range from two to one million. The oligomers are synthesized at designated areas on a substrate using a light-directed chemical process, wherein the substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support as described above.

[00228] In another aspect, an oligonucleotide may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application W095/251116 (Baldeschweiler et al.) which is incorporated herein in its entirety by reference.

[00229] A gene expression profile comprises the expression of a plurality of transcripts as measured by after hybridization with a sample. The transcripts of the invention may be used as elements on an array to produce a gene expression profile. In one embodiment, the array is used to diagnose or monitor the progression of disease. Researchers can assess and catalog the differences in gene expression between healthy and diseased tissues or cells.

[00230] For example, the transcript or probe may be labeled by standard methods and added to a biological sample from a patient under conditions for the formation of hybridization complexes. After an incubation period, the sample is washed and the amount of label (or signal) associated with hybridization complexes, is quantified and compared with a standard value. If complex formation in the patient sample is significantly altered (higher or lower) in comparison to either a normal or disease standard, then differential expression indicates the presence of a disorder.

[00231] In order to provide standards for establishing differential expression, normal and disease expression profiles are established. This is accomplished by combining a

sample taken from normal subjects, either animal or human or nonmammal, with a transcript under conditions for hybridization to occur. Standard hybridization complexes may be quantified by comparing the values obtained using normal subjects with values from an experiment in which a known amount of a purified sequence is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who were diagnosed with a particular condition, disease, or disorder. Deviation from standard values toward those associated with a particular disorder is used to diagnose that disorder.

[00232] By analyzing changes in patterns of gene expression, disease can be diagnosed at earlier stages before the patient is symptomatic. The invention can be used to formulate a prognosis and to design a treatment regimen. The invention can also be used to monitor the efficacy of treatment. For treatments with known side effects, the array is employed to improve the treatment regimen. A dosage is established that causes a change in genetic expression patterns indicative of successful treatment. Expression patterns associated with the onset of undesirable side effects are avoided.

[00233] In another embodiment, animal models which mimic a human disease can be used to characterize expression profiles associated with a particular condition, disease, or disorder; or treatment of the condition, disease, or disorder. Novel treatment regimens may be tested in these animal models using arrays to establish and then follow expression profiles over time. In addition, arrays may be used with cell cultures or tissues removed from animal models to rapidly screen large numbers of candidate drug molecules, looking for ones that produce an expression profile similar to those of known therapeutic drugs, with the expectation that molecules with the same expression profile will likely have similar therapeutic effects. Thus, the invention provides the means to rapidly determine the molecular mode of action of a drug.

[00234] Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies or in clinical trials or to monitor the treatment of an individual patient. Once the presence of a condition is established and a treatment protocol is initiated, diagnostic assays may be repeated on a regular basis to

determine if the level of expression in the patient begins to approximate that which is observed in a normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to years.

WORKING EXAMPLES:

1. Pancreatic cell line model system

[00235] Analysis of gene expression in various pancreatic cancer cell lines as well as pancreatic duct epithelial tissue has shown that the cell line Hs766T correlates well with normal tissue. For this reason, this cell line is reported in the literature as being a good surrogate for normal tissue in analyses of differential expression between pancreatic adenocarcinoma (and derived tumor lines) and normal tissue (or surrogate, Hs766T). The model system employed here involves the use of Hs766T as a “normal” reference to which cell surface expression in tumor derived cell lines is compared.

[00236] Differentially expressed E-cadherin and candidate modulators are validated in various tissues, cancer and normal pancreas and cell lines, to confirm that they are differentially expressed. Details of the pancreatic tumor cell lines that are used for this study, as well as the pancreatic line Hs766T are provided in Table 1 below.

Table 1. Cell Lines and Media

Cell line	ATCC Reference	Base medium	Non-essential			Sodium		Fetal Bovine Serum	
			Glutamine	amino acids	Carbonate	Pyruvate	Hepes		
Panc-1	CRL-1469	DMEM	2 mM	1% (w/v)	0.1% (w/v)	1 mM			10% (v/v)
Hs766t	HTB-134	DMEM	2 mM	1% (w/v)	0.1% (w/v)	1 mM			10% (v/v)
SU.86.86	CRL-1837	DMEM	2 mM	1% (w/v)	0.1% (w/v)	1 mM			10% (v/v)
AsPC1	CRL-1682	RPMI	2 mM	1% (w/v)	0.1% (w/v)	1 mM	10 mM		20% (v/v)
HPAF II	CRL-1997	DMEM	2 mM	1% (w/v)	0.1% (w/v)	1 mM			10% (v/v)
HPAC	CRL-2119	DMEM	2 mM	1% (w/v)	0.1% (w/v)	1 mM			10% (v/v)
Mia-Paca-2	CRL-1420	DMEM	2 mM	1% (w/v)	0.1% (w/v)	1 mM			10% (v/v)
Mpanc-96	CRL-2380	RPMI	2 mM	1% (w/v)	0.1% (w/v)	1 mM	10 mM		10% (v/v)
BxPC-3	CRL-1687	RPMI	2 mM	1% (w/v)	0.1% (w/v)	1 mM	10 mM		10% (v/v)
Capan-2	HTB-80	DMEM	2 mM	1% (w/v)	0.1% (w/v)	1 mM			10% (v/v)

2. Pancreatic cancer cell line culture

[00237] Cell lines are grown in a culturing medium that is supplemented as necessary with growth factors and serum (as described in Table 1). Cultures are established from frozen stocks in which the cells are suspended in a freezing medium (cell culture medium with 10% DMSO [v/v]) and flash frozen in liquid nitrogen. Frozen stocks prepared this way are stored in liquid nitrogen vapor. Cell cultures are established by rapidly thawing frozen stocks at 37°C. Thawed stock cultures are slowly transferred to a culture vessel containing a large volume of culture medium that is supplemented. For maintenance of culture, cells are seeded at 1×10^5 cells/per ml in a suitable medium and incubated at 37°C until confluence of cells in the culture vessel exceeds 50% by area. At this time, cells are harvested from the culture vessel using enzymes or EDTA where necessary. The density of harvested, viable cells is estimated by hemocytometry and the culture reseeded as above. A passage of this nature is repeated no more than 25 times at which point the culture is destroyed and reestablished from frozen stocks as described above.

[00238] For analyses of cell surface protein expression in cultured cell lines, cells are grown as described above. At a period 24 h prior to the experiment, the cell line is passaged as described above. This yielded cell densities that are <50% confluent and growing exponentially. Typically, triplicate analyses of differential expression are performed for each line relative to Hs766T for the purpose of identifying statistically significant reproducible differentially expressed proteins.

3. Antibody Development

Polyclonal antibody preparations:

[00239] Polyclonal antibodies against recombinant proteins are raised in rabbits (Green Mountain Antibodies, Burlington, VT). Briefly, two New Zealand rabbits are immunized with 0.1mg of antigen in complete Freund's adjuvant. Subsequent immunizations are carried out using 0.05mg of antigen in incomplete Freund's adjuvant at days 14, 21 and 49. Bleeds are collected and screened for recognition of the antigen by solid phase ELISA and western blot analysis. The IgG fraction is separated by

centrifugation at 20,000Xg for 20 minutes followed by a 50% ammonium sulfate cut. The pelleted protein is resuspended in 5mM Tris and separated by ion exchange chromatography. Fractions are pooled based on IgG content. Antigen-specific antibody is affinity purified using Pierce AMINOLINK resin coupled to the appropriate antigen.

Isolation of Antibody Fragments Directed Against E-cadherin from A Library of scFvs

[00240] Naturally occurring V-genes isolated from human PBLs are constructed into a library of antibody fragments which contain reactivities against E-cadherin to which the donor may or may not have been exposed (see e.g., U.S. Pat. No. 5,885,793 incorporated herein by reference in its entirety).

[00241] Rescue of the Library: A library of scFvs is constructed from the RNA of human PBLs as described in PCT publication WO 92/01047. To rescue phage displaying antibody fragments, approximately 10^9 *E. coli* harboring the phagemid are used to inoculate 50 ml of 2x TY containing 1% glucose and 100 µg/ml of ampicillin (2.times.TY-AMP-GLU) and grown to an O.D. of 0.8 with shaking. Five ml of this culture is used to inoculate 50 ml of 2.times.TY-AMP-GLU, 2×10^8 TU of delta gene 3 helper (M13 delta gene III, see PCT publication WO 92/01047) are added and the culture incubated at 37 °C. for 45 minutes without shaking and then at 37 °C. for 45 minutes with shaking. The culture is centrifuged at 4000 r.p.m. for 10 min. and the pellet resuspended in 2 liters of 2 x TY containing 100 µg/ml ampicillin and 50 µg/ml kanamycin and grown overnight. Phage are prepared as described in PCT publication WO 92/01047.

[00242] M13 delta gene III is prepared as follows: M13 delta gene III helper phage does not encode gene III protein, hence the phage(mid) displaying antibody fragments have a greater avidity of binding to antigen. Infectious M13 delta gene III particles are made by growing the helper phage in cells harboring a pUC19 derivative supplying the wild type gene III protein during phage morphogenesis. The culture is incubated for 1 hour at 37 °C without shaking and then for a further hour at 37 °C with shaking. Cells are spun down (IEC-CENTRA 8,400 r.p.m. for 10 min), resuspended in 300 ml 2 x TY broth containing 100 µg ampicillin/ml and 25 µg kanamycin/ml (2 x TY-AMP-KAN) and grown overnight, shaking at 37 °C. Phage particles are purified and concentrated from the

culture medium by two PEG-precipitations (Sambrook et al., 1990), resuspended in 2 ml PBS and passed through a 0.45 μm filter (MINISART NML; Sartorius) to give a final concentration of approximately 1013 transducing units/ml (ampicillin-resistant clones).

[00243] Panning of the Library: IMMUNOTUBES (Nunc) are coated overnight in PBS with 4 ml of either 100 $\mu\text{g/ml}$ or 10 $\mu\text{g/ml}$ of a polypeptide of the present invention. Tubes are blocked with 2% Marvel-PBS for 2 hours at 37 °C and then washed 3 times in PBS. Approximately 1013 TU of phage is applied to the tube and incubated for 30 minutes at room temperature tumbling on an over and under turntable and then left to stand for another 1.5 hours. Tubes are washed 10 times with PBS 0.1% Tween-20 and 10 times with PBS. Phage are eluted by adding 1 ml of 100 mM triethylamine and rotating 15 minutes on an under and over turntable after which the solution is immediately neutralized with 0.5 ml of 1.0M Tris-HCl, pH 7.4. Phages are then used to infect 10 ml of mid-log *E. coli* TG1 by incubating eluted phage with bacteria for 30 minutes at 37 °C. The *E. coli* are then plated on TYE plates containing 1% glucose and 100 $\mu\text{g/ml}$ ampicillin. The resulting bacterial library is then rescued with delta gene 3 helper phage as described above to prepare phage for a subsequent round of selection. This process is then repeated for a total of 4 rounds of affinity purification with tube-washing increased to 20 times with PBS, 0.1% Tween-20 and 20 times with PBS for rounds 3 and 4.

[00244] Characterization of Binders: Eluted phage from the 3rd and 4th rounds of selection are used to infect *E. coli* HB 2151 and soluble scFv is produced (Marks, et al., 1991) from single colonies for assay. ELISAs are performed with microtitre plates coated with either 10 $\mu\text{g/ml}$ of the polypeptide of the present invention in 50 mM bicarbonate pH 9.6. Clones positive in ELISA are further characterized by PCR fingerprinting (see, e.g., PCT publication WO 92/01047) and then by sequencing.

Monoclonal Antibody Generation

[00245] i) Materials:

[00246] 1) Complete Media No Sera (CMNS) for washing of the myeloma and spleen cells; Hybridoma medium CM - HAT {Cell Mab (BD), 10% FBS (or HS); 5% Origen

HCF (hybridoma cloning factor) containing 4mM L-glutamine and antibiotics} to be used for plating hybridomas after the fusion.

[00247] 2) Hybridoma medium CM - HT (NO AMINOPTERIN) (Cell Mab (BD), 10% FBS 5% Origen HCF containing 4mM L-glutamine and antibiotics) to be used for fusion maintenance are stored in the refrigerator at 4-6 °C. The fusions are fed on days 4, 8, and 12, and subsequent passages. Inactivated and pre-filtered commercial Fetal Bovine serum (FBS) or Horse Serum (HS) are thawed and stored in the refrigerator at 4 °C and must be pretested for myeloma growth from single cells.

[00248] 3) The L-glutamine (200mM, 100X solution), which is stored at -20 °C freezer, is thawed and warmed until completely in solution. The L-glutamine is dispensed into media to supplement growth. L-glutamine is added to 2 mM for myelomas, and 4 mM for hybridoma media. Further the Penicillin, Streptomycin, Amphotericin (antibacterial-antifungal stored at -20 °C) is thawed and added to Cell Mab Media to 1%.

[00249] 4) Myeloma growth media is Cell Mab Media (Cell Mab Media, QUANTUM YIELD from BD is stored in the refrigerator at 4 °C in the dark) which are added L-glutamine to 2 mM and antibiotic/antimycotic solution to 1% and is called CMNS.

[00250] 5) 1 bottle of PEG 1500 in Hepes (Roche, NJ)

[00251] 6) 8-Azaguanine is stored as the dried powder supplied by SIGMA at -700 °C until needed. Reconstitute 1 vial / 500 ml of media and add entire contents to 500 ml media (eg. 2 vials/ liter).

[00252] 7) Myeloma Media is CM which has 10% FBS (or HS) and 8-Aza (1 X) stored in the refrigerator at 4 °C.

[00253] 8) Clonal cell medium D (Stemcell, Vancouver) contains HAT and methyl cellulose for semi-solid direct cloning from the fusion.

[00254] 9) Hybridoma supplements HT [hypoxanthine, thymidine] are to be used in medium for the selection of hybridomas and maintenance of hybridomas through the cloning stages respectively.

[00255] 10) Origen HCF can be obtained directly from Igen and is a cell supernatant produced from a macrophage-like cell-line. It can be thawed and aliquoted to 15 ml tubes at 5 ml per tube and stored frozen at -20 °C. Positive Hybridomas are fed HCF through the first subcloning and are gradually weaned. It is not necessary to continue to supplement unless you have a particularly difficult hybridoma clone. This and other additives have been shown to be more effective in promoting new hybridoma growth than conventional feeder layers.

[00256] ii) Procedure

[00257] To generate monoclonal antibodies, mice are immunized with 5-50 ug of antigen either intra-peritoneal (i.p.) or by intravenous injection in the tail vein (i.v.). Typically, the antigen used is a recombinant protein that is generated as described above. The primary immunization takes place 2 months prior to the harvesting of splenocytes from the mouse and the immunization is typically boosted by i.v. injection of 5-50 ug of antigen every two weeks. At least one week prior to expected fusion date, a fresh vial of myeloma cells is thawed and cultured. Several flasks at different densities are maintained in order that a culture at the optimum density is ensured at the time of fusion. The optimum density is determined to be $3-6 \times 10^5$ cells/ml. Two to five days before the scheduled fusion, a final immunization is administered of ~5ug of antigen in PBS i.p. or i.v.

[00258] Myeloma cells are washed with 30 ml serum free media by centrifugation at $500 \times g$ at 4°C for 5 minutes. Viable cell density is determined in resuspended cells using hemocytometry and vital stains. Cells resuspended in complete growth medium are stored at 37°C during the preparation of splenocytes. Meanwhile, to test aminopterin sensitivity, 1×10^6 myeloma cells are transferred to a 15ml conical tube and centrifuged at 500g at 4°C for 5 minutes. The resulting pellet is resuspended in 15 ml of HAT media and cells plated at 2 drops/well on a 96 well plate.

[00259] To prepare splenocytes from immunized mice, the animals are euthanised and submerged in 70% ETOH. Under sterile conditions, the spleen is surgically removed and placed in 10ml of RPMI medium supplemented with 20% fetal calf serum in a Petri dish.

Cells are extricated from the spleen by infusing the organ with medium >50 times using a 21g syringe.

[00260] Cells are harvested and washed by centrifugation (at 500g at 4°C for 5 minutes) with 30 ml of medium. Cells are resuspended in 10 ml of medium and the density of viable cells determined by hemocytometry using vital stains. The splenocytes are mixed with myeloma cells at a ratio of 5:1 (spleen cells: myeloma cells). Both the myeloma and spleen cells are washed 2 more times with 30 ml of RPMI-CMNS. Spin at 800 rpm for 12 minutes.

[00261] Supernatant is removed and cells are resuspended in 5 ml of RPMI-CMNS and are pooled to bring the volume to 30 ml and spun down as before. The cell pellet is broken up by gentle tapping and resuspended in 1 ml of BMB PEG1500 (prewarmed to 37°C) added dropwise with a 1 cc needle over 1 minute..

[00262] RPMI-CMNS is added to the PEG cells to slowly dilute out the PEG. Cells are centrifuged and diluted in 5 ml of Complete media and 95 ml of Clonacell Medium D (HAT) media (with 5 ml of HCF). The cells are plated out at 10 ml per small petri plate.

[00263] Myeloma/HAT control is prepared as follows. Dilute about 1000 P3X63 Ag8.653 myeloma cells into 1 ml of medium D and transfer into a single well of a 24 well plate. Plates are placed in incubator, with two plates inside of a large petri plate, with an additional petri plate full of distilled water, for 10-18 days under 5% CO₂ overlay at 37°C. Clones are picked from semisolid agarose into 96 well plates containing 150-200 ul of CM- HT. Supernatants are screened 4 days later in ELISA, and positive clones are moved up to 24 well plates. Heavy growth will require changing of the media at day 8 (+/- 150 ml). One should further decrease the HCF to 0.5% (gradually- 2%, then 1%, then 0.5%) in the cloning plates.

[00264] For further references see Kohler G, and C. Milstein Continuous cultures of fused cells secreting antibody of predefined specificity.1975. Nature 256: 495-497; Lane, R.D. A short duration polyethylene glycol fusion technique for increasing production of

monoclonal antibody-secreting hybridomas. 1985. J. Immunol. Meth. 81:223-228; Harlow, E. and D. Lane. Antibodies: A Laboratory Manual. Cold Spring Harbor Laboratory Press. 1988; Kubitz, D. The Scripps Research Institute. La Jolla. Personal Communication; Zhong, G., Berry, J.D., and Choukri, S. (1996) Mapping epitopes of Chlamydia trachomatis neutralizing monoclonal antibodies using phage random peptide libraries. J. Indust. Microbiol. Biotech. 19, 71-76; Berry, J.D. , Licea, A., Popkov, M., Cortez, X., Fuller, R., Elia, M., Kerwin, L., and C.F. Barbas III. (2003) Rapid monoclonal antibody generation via dendritic cell targeting in vivo. Hybridoma and Hybridomics 22 (1), 23-31.

4. Expression Validation

mRNA Expression Validation by TAQMAN

[00265] Expression of mRNA is quantitated by RT-PCR using TAQMAN technology. The TAQMAN system couples a 5' fluorogenic nuclease assay with PCR for real time quantitation. A probe is used to monitor the formation of the amplification product.

[00266] Total RNA is isolated from cancer model cell lines using the RNEASY kit (Qiagen) per manufacturer's instructions and included DNase treatment. Normal human tissue RNAs are acquired from commercial vendors (Ambion, Austin, TX; Stratagene, La Jolla, CA, BioChain Institute, Newington, NH) as are RNAs from matched disease/normal tissues.

[00267] Target transcript sequences are identified for the differentially expressed peptides by searching the BlastP database. TAQMAN assays (PCR primer/probe set) specific for those transcripts are identified by searching the CELERA DISCOVERY SYSTEM (CDS) database. The assays are designed to span exon-exon borders and do not amplify genomic DNA.

[00268] The TAQMAN primers and probe sequences are designed by Applied Biosystems (AB) as part of the ASSAYS ON DEMAND product line or by custom design through the AB ASSAYS BY DESIGN service.

[00269] RT-PCR is accomplished using AMPLITAQGOLD and MULTISCRIBE reverse transcriptase in the ONE STEP RT-PCR Master Mix reagent kit (AB) according to the manufacturer's instructions. Probe and primer concentrations are 250 nM and 900 nM, respectively, in a 15 μ l reaction. For each experiment, a master mix of the above components is made and aliquoted into each optical reaction well. Eight nanograms of total RNA is used as the template. Each sample is assayed in triplicate. Quantitative RT-PCR is performed using the ABI PRISM 7900HT SEQUENCE DETECTION SYSTEM (SDS). Cycling parameters follow: 48°C for 30 min. for one cycle; 95 °C for 10 min for one cycle; 95 °C for 15 sec, 60 °C for 1 min. for 40 cycles.

[00270] The SDS software calculates the threshold cycle (C_T) for each reaction, and C_T values are used to quantitate the relative amount of starting template in the reaction. The C_T values for each set of three reactions are averaged for all subsequent calculations.

[00271] Data are analyzed for differences in expression using an endogenous control for normalization, and measuring expression relative to a normal tissue or normal cell line reference. The choice of endogenous control is determined empirically by testing various candidates against the cell line and tissue RNA panels and selecting the one with the least variation in expression. Relative changes in expression are quantitated using the $2^{-\Delta\Delta C_T}$ Method. (See Livak, *et al.*, 2001, Methods 25: 402-408; User bulletin #2: ABI PRISM 7700 SEQUENCE DETECTION SYSTEM.)

Protein Expression Validation by Western

[00272] Western blot analysis of target proteins is carried out using whole cell extracts prepared from each of the pancreatic cell lines. To make cell extracts, the cells are resuspended in Lysis buffer (125mM Tris, pH 7.5, 150mM NaCl, 2% SDS, 5mM EDTA, 0.5% NP-40) and passed through a 20-gauge needle. Lysates are centrifuged at 5,000Xg for 5 minutes at 4°C. The supernatants are collected and a protease inhibitor cocktail (Sigma) is added. The Pierce BCA assay is used to quantitate total protein. Samples are separated by SDS-PAGE and transferred to either a nitrocellulose or PVDF membrane. The WESTERN BREEZE kit from Invitrogen is used for western blot analysis. Primary antibodies are either purchased from commercially available sources or prepared using

one of the methods described in Section 3. For this application, antibodies are typically diluted 1:500 to 1:10,000 in a diluent buffer. Blots are developed using Pierce NBT.

Tissue Flow Cytometry Analysis

[00273] Post tissue processing, cells are sorted by flow cytometry known in the art to enrich for epithelial cells. Alternatively, cells isolated from pancreatic tissue are stained directly with EpCAM (for epithelial cells) and the specific antibody for E-cadherin. Cell numbers and viability are determined by PI exclusion (GUAVA) for cells isolated from both normal and tumor pancreatic tissue. A minimum of 0.5×10^6 cells are used for each analysis. Cells are washed once with Flow Staining Buffer (0.5% BSA, 0.05% NaN₃ in D-PBS).

[00274] To the cells, 20 μ l of an antibody against E-cadherin are added. An additional 5 μ l of EpCAM antibody conjugated to APC are added when unsorted cells are used in the experiment. Cells are incubated with antibodies for 30 minutes at 4°C. Cells are washed once with Flow Staining Buffer and either analyzed immediately on the LSR flow cytometry apparatus or fixed in 1% formaldehyde and store at 4°C until LSR analysis.

5. Detection and Diagnosis of E-cadherin by Liquid Chromatography and Mass Spectrometry (LC/MS)

[00275] The differential expression of proteins in disease and healthy samples are quantitated using Mass Spectrometry and ICAT (Isotope Coded Affinity Tag) labeling. ICAT is an isotope label technique that allows for discrimination between two populations of proteins, such as from a healthy and a disease sample that are pooled together for experimental purposes or two acquisitions of the same sample for classification of true sample peptides from LC/MS noise artifacts.

[00276] The proteins from cells are prepared by methods known in the art. The LC/MS spectra are collected for the labeled samples and processed using the following steps:

[00277] The raw scans from the LC/MS instrument are subjected to peak detection and noise reduction using standard software. Filtered peak lists are then used to detect

“features” corresponding to specific peptides from the original sample(s). Features are characterized by their mass/charge, charge, retention time, isotope pattern and intensity.

[00278] Similar experiments are repeated in order to increase the confidence in detection of a peptide. These multiple acquisitions are computationally aggregated into one experiment. Experiments involving healthy and disease samples use the known effects of the ICAT label to classify the peptides as originating from a particular sample or from both samples. The intensity of a peptide present in both healthy and disease samples is used to calculate the differential expression, or relative abundance, of the peptide. The intensity of a peptide found exclusively in one sample is used to calculate a theoretical expression ratio for that peptide (singleton). Expression ratios are calculated for each peptide of each replicate of the experiment.

[00279] Statistical tests are performed to assess the robustness of the data and statistically significant differentials selected. These tests a) ensure that similar features are detected in all replicates of the experiment; b) assess the distribution of the log ratios of all peptides (a Gaussian is expected); c) calculate the overall pair wise correlations between ICAT LC/MS maps to ensure that the expression ratios for peptides are reproducible across the multiple replicates; and d) aggregate multiple experiments in order to compare the expression ratio of a peptide in multiple diseases or disease samples.

Results

[00280] Peptides corresponding to E-cadherin proteins (SEQ ID NOs: 1, 2 and 3) are overexpressed in the cell lines HPAF II, Capan-2, BxPC-3, and SU.86.86.

8. Expression Validation by IHC in tissue sections

Tissue sections

[00281] Paraffin embedded, fixed tissue sections are obtained from a panel of normal tissues (Adrenal, Bladder, Lymphocytes, Bone Marrow, Breast, Cerebellum, Cerebral cortex, Colon, Endothelium, Eye, Fallopian tube, Small Intestine, Heart, Kidney (glomerulus, tubule), Liver, Lung, Testes and Thyroid) as well as 30 tumor samples with matched normal adjacent tissues from pancreas, lung, colon, prostate, ovarian and breast. In addition, other tissues are selected for testing such as bladder renal, hepatocellular, pharyngeal and gastric tumor tissues.

[00282] Esophageal replicate sections are also obtained from numerous tumor types (Bladder Cancer, Lung Cancer, Breast Cancer, Melanoma, Colon Cancer, Non-Hodgkins Lymphoma, Endometrial Cancer, Ovarian Cancer, Head and Neck Cancer, Prostate Cancer, Leukemia [ALL and CML] and Rectal Cancer). Sections are stained with hemotoxylin and eosin and histologically examined to ensure adequate representation of cell types in each tissue section.

[00283] An identical set of tissues are obtained from frozen sections and are used in those instances where it is not possible to generate antibodies that are suitable for fixed sections. Frozen tissues do not require an antigen retrieval step.

Hemotoxylin and Eosin staining of paraffin embedded, fixed tissue sections.

[00284] Sections are deparaffinized in 3 changes of xylene or xylene substitute for 2-5 minutes each. Sections are rinsed in 2 changes of absolute alcohol for 1-2 minutes each, in 95% alcohol for 1 minute, followed by 80% alcohol for 1 minute. Slides are washed well in running water and stained in Gill solution 3 hemotoxylin for 3 to 5 minutes. Following a vigorous wash in running water for 1 minute, sections are stained in Scott's solution for 2 minutes. Sections are washed for 1 min in running water then counterstained in Eosin solution for 2-3 minutes depending upon development of desired staining intensity. Following a brief wash in 95% alcohol, sections are dehydrated in three changes of absolute alcohol for 1 minute each and three changes of xylene or xylene substitute for 1-2 minutes each. Slides are coverslipped and stored for analysis.

Optimization of antibody staining

[00285] For each antibody, a positive and negative control sample is generated using data from the ICAT analysis of the pancreatic cancer cell lines. Cell lines are selected that are known to express low levels of a particular target as determined from the ICAT data. This cell line is the reference normal control "Hs766T." Similarly, a pancreatic tumor line known to overexpress the target is selected as positive control.

Antigen retrieval

[00286] Sections are deparaffinized and rehydrated by washing 3 times for 5 minutes in xylene; two times for 5 minutes in 100% ethanol; two times for 5 minutes in 95% ethanol; and once for 5 minutes in 80% ethanol. Sections are then placed in endogenous blocking solution (methanol + 2% hydrogen peroxide) and incubated for 20 minutes at room temperature. Sections are rinsed twice for 5 minutes each in deionized water and twice for 5 minutes in phosphate buffered saline (PBS), pH 7.4. Alternatively, where necessary sections are deparaffinized by High Energy Antigen Retrieval as follows: sections are washed three times for 5 minutes in xylene; two times for 5 minutes in 100% ethanol; two times for 5 minutes in 95% ethanol; and once for 5 minutes in 80% ethanol. Sections are placed in a Coplin jar with dilute antigen retrieval solution (10 mM citrate acid, pH 6). The Coplin jar containing slides is placed in a vessel filled with water and microwaved on high for 2 - 3 minutes (700 watt oven). Following cooling for 2 - 3 minutes, steps 3 and 4 are repeated four times (depending on the tissue), followed by cooling for 20 minutes at room temperature. Sections are then rinsed in deionized water, two times for 5 minutes, placed in modified endogenous oxidation blocking solution (PBS + 2% hydrogen peroxide) and rinsed for 5 minutes in PBS.

Blocking and Staining

[00287] Sections are blocked with PBS/1% bovine serum albumin (PBA) for 1 hour at room temperature followed by incubation in normal serum diluted in PBA (2%) for 30 minutes at room temperature to reduce non-specific binding of antibody. Incubations are performed in a sealed humidity chamber to prevent air-drying of the tissue sections. (The

choice of blocking serum is the same as the species of the biotinylated secondary antibody). Excess antibody is gently removed by shaking and sections covered with primary antibody diluted in PBA and incubated either at room temperature for 1 hour or overnight at 4°C. (Care is taken that the sections do not touch during incubation). Sections are rinsed twice for 5 minutes in PBS, shaking gently. Excess PBS is removed by gently shaking. The sections are covered with diluted biotinylated secondary antibody in PBA and incubated for 30 minutes to 1 hour at room temperature in the humidity chamber. If using a monoclonal primary antibody, addition of 2% rat serum is used to decrease the background on rat tissue sections. Following incubation, sections are rinsed twice for 5 minutes in PBS, shaking gently. Excess PBS is removed and sections incubated for 1 hour at room temperature in VECTASTAIN ABC reagent (Vector Laboratories, Burlingame, CA) according to kit instructions. The lid of the humidity chamber is secured during all incubations to ensure a moist environment. Sections are rinsed twice for 5 minutes in PBS, shaking gently.

Develop and Counterstain

[00288] Sections are incubated for 2 minutes in peroxidase substrate solution that is made up immediately prior to use as follows: 10mg diaminobenzidine (DAB) dissolved in 10ml 50 mM sodium phosphate buffer, pH 7.4; 12.5 microliters 3% $\text{CoCl}_2/\text{NiCl}_2$ in deionized water; 1.25 microliters hydrogen peroxide.

[00289] Slides are rinsed well three times for 10 min in deionized water and counterstained with 0.01% Light Green acidified with 0.01% acetic acid for 1-2 minutes depending on intensity of counterstain desired.

[00290] Slides are rinsed three times for 5 minutes with deionized water and dehydrated two times for 2 minutes in 95% ethanol; two times for 2 minutes in 100% ethanol; and two times for 2 minutes in xylene. Stained slides are mounted for visualization by microscopy.

7. IHC Staining of Frozen Tissue Sections

[00291] Fresh tissues are embedded carefully in OCT in a plastic mold, without trapping air bubbles surrounding the tissue. Tissues are frozen by setting the mold on top of liquid nitrogen until 70-80% of the block turns white at which point the mold is placed on dry ice. The frozen blocks are stored at -80°C. Blocks are sectioned with a cryostat with care taken to avoid warming to greater than -10°C. Initially, the block is equilibrated in the cryostat for about 5 minutes and 6-10mm sections are cut sequentially. Sections are allowed to dry for at least 30 minutes at room temperature. Following drying, tissues are stored at 4°C for short term and -80°C for long term storage.

[00292] Sections are fixed by immersing in acetone jar for 1-2 minutes at room temperature, followed by drying at room temperature. Primary antibody is added (diluted in 0.05 M Tris-saline [0.05 M Tris, 0.15 M NaCl, pH 7.4], 2.5% serum) directly to the sections by covering the section dropwise to cover the tissue entirely. Binding is carried out by incubation a chamber for 1 hour at room temperature. Without letting the sections dry out, the secondary antibody (diluted in Tris-saline/2.5% serum) is added in a similar manner to the primary and incubated as before (at least 45 minutes). Following incubation, the sections are washed gently in Tris-saline for 3-5 minutes and then in Tris-saline/2.5% serum for another 3-5 minutes. If a biotinylated primary antibody is used, in place of the secondary antibody incubation, slides are covered with 100 ul of diluted alkaline phosphatase conjugated streptavidin, incubated for 30 minutes at room temperature and washed as above. Sections are incubated with alkaline phosphatase substrate (1mg/ml Fast Violet; 0.2 mg/ml Naphthol AS-MX phosphate in Tris-Saline pH 8.5) for 10-20 minutes until the desired positive staining is achieved at which point the reaction is stopped by washing twice with Tris-saline. Slides are counter-stained with Mayer's hematoxylin for 30 seconds and washed with tap water for 2-5 minutes. Sections are mounted with Mount coverslips and mounting media.

8. Assay for Antibody Dependent Cellular Cytotoxicity (AOCC)

[00293] Cultured tumor cells are labeled with 100 μCi ^{51}Cr for 1 hour (*see* Livingston, *et al.*, 1997, Cancer Immunol. Immunother. 43; 324-330). After being washed three times

with culture medium, cells are resuspended at 10^5 /ml, and 100 μ l/well are plated onto 96-well round-bottom plates. A range of antibody concentrations are applied to the wells, including an isotype control together with donor peripheral blood mononuclear cells that are plated at a 100:1 and 50:1 ratio. After an 18-h incubation at 37°C, supernatant (30 μ l/well) is harvested and transferred onto LUMAPLATE 96 (Packard), dried, and read in a Packard TOP-COUNT NXT γ counter. Each measurement is carried out in triplicate. Spontaneous release is determined by cpm of tumor cells incubated with medium and maximum release by cpm of tumor cells plus 1% Triton X-100 (Sigma). Specific lysis is defined as: % specific lysis = [(experimental release–spontaneous release)/(maximum release–spontaneous release)] $\times$ 100. The percent ADCC is expressed as peak specific lysis postimmune subtracted by preimmune percent specific lysis. A doubling of the ADCC to >20% is considered significant.

9. Assay for Complement Dependent Cytotoxicity (CDC)

[00294] Chromium release assays to assess complement-mediated cytotoxicity are performed for each patient at various time points (Dickler, *et al.*, 1999, Clin. Cancer Res. 5, 2773-2779). Cultured tumor cells are washed in FCS-free media two times, resuspended in 500 μ l of media, and incubated with 100 μ Ci 51 Cr per 10 million cells for 2 h at 37°C. The cells are then shaken every 15 min for 2 h, washed 3 times in media to achieve a concentration of approximately 20,000 cells/well, and then plated in round-bottom plates. The plates contain either 50 μ l cells plus 50 μ l monoclonal antibody, 50 μ l cells plus serum (pre- and post-therapy), or 50 μ l cells plus mouse serum as a control. The plates are incubated in a cold room on a shaker for 45 min. Human complement of a 1:5 dilution (resuspended in 1 ml of ice-cold water and diluted with 3% human serum albumin) is added to each well at a volume of 100 μ l. Control wells include those for maximum release of isotope in 10% Triton X-100 (Sigma) and for spontaneous release in the absence of complement with medium alone. The plates are incubated for 2 h at 37°C, centrifuged for 3 min, and then 100 μ l of supernatant is removed for radioactivity counting. The percentage of specific lysis is calculated as follows: % cytotoxicity = [(experimental release–spontaneous release)/(maximum release–spontaneous release)] $\times$ 100. A doubling of the CDC to >20% is considered significant.

10. In vitro Assays in Cell Lines

[00295] LIPOFECTAMINE is purchased from Invitrogen (Carlsbad, CA) and GENESILENCER from Gene Therapy Systems (San Diego, CA). Synthetic siRNA oligonucleotides are from Dharmacon (Lafayette, CO), Qiagen (Valencia, CA) or Ambion (Austin, TX). RNEASY 96 Kit is purchased from Qiagen (Valencia, CA). APOP-ONE homogeneous caspase-3/7 kit and CELLTITER 96 Aqueous One Solution Cell Proliferation Assay are both purchased from Promega (Madison, WI). Cell invasion assay kits from purchased from Chemicon (Temecula, CA). RIBOGREEN RNA Quantitation Kit is purchased from Molecular probes (Eugene, OR).

RNAi

[00296] RNAi is performed by using SMARTPOOLS (Dharmacon), 4-FOR SILENCING siRNA duplexes (Qiagen) or scrambled negative control siRNA (Ambion). Transient transfections are carried out in triplicate by using either LIPOFECTAMINE 2000 from Invitrogen (Carlsbad, CA) or by using GENESILENCER from Gene Therapy Systems (San Diego, CA) in methods described below. One to four days after transfections, total RNA is isolated using the RNEASY 96 Kit (Qiagen) according to manufacturer's instructions and expression of mRNA is quantitated using the TAQMAN technology. Protein expression levels are examined by flow cytometry. Apoptosis and proliferation assays are performed daily using APOP-ONE homogeneous caspase-3/7 kit and CELLTITER 96 Aqueous One Solution Cell Proliferation Assay.

[00297] Transient transfections are carried out on sub-confluent pancreatic cancer cell lines as previously described. Elbashir, S.M. et al. (2001) Nature 411: 494-498; Caplen, N.J. et al. (2001) Proc Natl Acad Sci USA 98: 9742-9747; Sharp, P.A. (2001) Genes and Development 15: 485-490. Synthetic siRNA to gene of interest or scrambled negative control siRNA is transfected using LIPOFECTAMINE according to manufacturer's instructions. Cells are plated in 96 well plates in antibiotic-free medium. The next day, the transfection reagent and siRNA are prepared for transfection as follows: Each 0.1-1ul of LIPOFECTAMINE 2000 and 10-150mM siRNA are resuspended 25ul serum-free media and incubated at room temperature for 5 minutes. After incubation, the diluted

siRNA and the LIPOFECTAMINE 2000 are combined and incubated for 20 minutes at room temperature. The cells are then washed and the combined siRNA-LIPOFECTAMINE 2000 reagent added. After further 4 hours incubation, 50 μ l serum containing medium is added to each well. One and four days after transfection, expression of mRNA is quantitated by RT-PCR using the TAQMAN technology and protein expression levels are examined by flow cytometry. Apoptosis and proliferation assays are performed daily using APOP-ONE homogeneous caspase-3/7 kit and CELLTITER 96 Aqueous One Solution Cell Proliferation Assay.

RNAi Transfections- GeneSilencer

[00298] Transient transfections are carried out on sub-confluent pancreatic cancer cell lines as previously described (Elbashir, *et al.*, 2001, Nature 411: 494-498; Caplen, *et al.*, 2001, Proc. Natl. Acad. Sci. USA 98: 9742-9747; Sharp, 2001, Genes and Development 15: 485-490). Synthetic siRNA to gene of interest or scrambled negative control siRNA is transfected using GENESILENCER according to manufacturer's instructions. Cells are plated in 96 well plates in antibiotic-free medium. The next day, the transfection reagent and the synthetic siRNA are prepared for transfection as follows: predetermined amount of GENE SILENCER is diluted in serum-free media to a final volume of 20 μ l per well. After resuspending 10-150 mM siRNA in 20 μ l serum-free media, the reagents are combined and incubated at room temperature for 5-20 minutes. After incubation, the siRNA-GENE SILENCER reagent is added to each well and incubated in a 37°C incubator for 4 hours before an equal volume of serum containing media is added back to the cultured cells. The cells are then incubated for 1 to 4 days before mRNA, protein expression and effects on apoptosis and proliferation are examined.

Testing of function blocking antibodies

[00299] Sub-confluent pancreatic cancer cell lines are serum-starved overnight. The next day, serum-containing media is added back to the cells in the presence of 5-50ng/ml of function blocking antibodies. After 2 or 5 days incubation at 37°C, 5% CO₂, antibody binding is examined by flow cytometry and apoptosis and proliferation are examined by using protocols described below.

[00300] Apoptosis assay is performed using the APOP-ONE homogeneous caspase-3/7 kit from Promega according to the manufacturer's instructions.

[00301] Cell proliferation assay is performed using the CELLTITER 96 Aqueous One Solution Cell Proliferation Assay kit from Promega. 20 μ l of CELLTITER 96 Aqueous One Solution is added to 100 μ l of culture medium. The plates are then incubated for 1-4 hours at 37°C in a humidified 5% CO₂ incubator. After incubation, the change in absorbance is read at 490 nm.

Cell invasion

[00302] Cell invasion assay is performed using the 96-well cell invasion assay kit available from Chemicon. After the cell invasion chamber plates are adjusted to room temperature, 100 μ l serum-free media is added to the interior of the inserts. 1-2 hours later, cell suspensions of 1×10^6 cells/ml are prepared. Media is then carefully removed from the inserts and 100 μ l of prepared cells are added into the insert along with 0 to 50 ng of function-blocking antibodies. The cells are pre-incubated for 15 minutes at 37°C before 150 μ l of media containing 10% FBS is added to the lower chamber. The cells are then incubated for 48 hours at 37°C. After incubation, the cells from the top side of the insert are discarded and the invasion chamber plates are then placed on a new 96-well feeder tray containing 150 μ l of pre-warmed cell detachment solution in the wells. The plates are incubated for 30 minutes at 37°C and are periodically shaken. Lysis buffer/ dye solution (4ul CYQUANT Dye/ 300 μ l 4X lysis buffer) is prepared and added to each well of dissociation buffer/ cells on feeder tray. The plates are incubated for 15 minutes at room temperature before 150 μ l is transferred to a new 96-well plate. Fluorescence of invading cells is then read at 480 nm excitation and 520 nm emission.

Receptor internalization

[00303] For quantification of receptor internalization, ELISA assays are performed essentially as described by Daunt *et al.*, 1997, Mol. Pharmacol. 51, 711-720. The cell lines are plated at 6×10^5 cells per in a 24-well tissue culture dishes that have previously been coated with 0.1mg/ml poly-L-lysine. The next day, the cells are washed once with

PBS and incubated in DMEM at 37°C for several minutes. The agonist to the cell surface target of interest is then added at a pre-determined concentration in prewarmed DMEM to the wells. The cells are then incubated for various times at 37°C and reactions are stopped by removing the media and fixing the cells in 3.7% formaldehyde/TBS for 5 min at room temperature. The cells are then washed three times with TBS and nonspecific binding blocked with TBS containing 1% BSA for 45 min at room temperature. The first antibody is added at a pre-determined dilution in TBS/BSA for 1 h at room temperature. Three washes with TBS followed, and cells are briefly reblocked for 15min at room temperature. Incubation with goat anti-mouse conjugated alkaline phosphatase (Bio-Rad) diluted 1:1000 in TBS/BSA is carried out for 1h at room temperature. The cells are washed three times with TBS and a colorimetric alkaline phosphatase substrate is added. When the adequate color change is reached, 100- μ l samples are taken for colorimetric readings.

mRNA expression

[00304] Expression of mRNA is quantitated by RT-PCR using TAQMAN technology. Total RNA is isolated from cancer model cell lines using the RNEASY 96 kit (Qiagen) per manufacturer's instructions and included DNase treatment. Target transcript sequences are identified for the differentially expressed peptides by searching the BlastP database. TAQMAN assays (PCR primer/probe set) specific for those transcripts are identified by searching the CELERA DISCOVERY SYSTEM (CDS) database. The assays are designed to span exon-exon borders and do not amplify genomic DNA. The TAQMAN primers and probe sequences are as designed by Applied Biosystems (AB) as part of the ASSAYS ON DEMAND product line or by custom design through the AB ASSAYS BY DESIGN service. RT-PCR is accomplished using AMPLITAQ GOLD and MULTISCRIBE reverse transcriptase in the ONE STEP RT-PCR Master Mix reagent kit (AB) according to the manufacturers instructions. Probe and primer concentrations are 900 nM and 250 nM, respectively, in a 25 μ l reaction. For each experiment, a master mix of the above components is made and aliquoted into each optical reaction well. 5ul of total RNA is the template. Each sample is assayed in triplicate. Quantitative RT-PCR is performed using the ABI PRISM 7900HT Sequence Detection System (SDS). Cycling

parameters follow: 48 °C for 30 min. for one cycle; 95 °C for 10 min for one cycle; 95 °C for 15 sec, 60 °C for 1 min. for 40 cycles.

[00305] The SDS software calculates the threshold cycle (C_T) for each reaction, and C_T values are used to quantitate the relative amount of starting template in the reaction. The C_T values for each set of three reactions are averaged for all subsequent calculations.

[00306] Total RNA is quantitated using the RIBOGREEN RNA Quantitation Kit according to manufacturer's instructions and the percent mRNA expression is calculated using total RNA for normalization. Percent knockdown is then calculated relative to the no- addition control.

[00307] Anti-E-cadherin antibodies inhibit cell proliferation, as measured using MTS (3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethanoxypheyl)-2-(4-sulfophenyl)-2H-tetrazolium) cell proliferation assays (see e.g. Emerman and Eaves, 1994, Bone Marrow Transplantation 13:285) (Figure 2).

11. In Vivo Studies Using Antibodies

Treatment of pancreatic cancer cells with monoclonal antibodies.

[00308] Pancreatic cancer cells are seeded at a density of 4×10^4 cells per well in 96-well microtiter plates and allowed to adhere for 2 hours. The cells are then treated with different concentrations of anti-E-cadherin monoclonal antibody (Mab) or irrelevant isotype matched (anti-rHuIFN- γ Mab) at 0.05, 0.5 or 5.0 μ g/ml. After a 72 hour incubation, the cell monolayers are stained with crystal violet dye for determination of relative percent viability (RPV) compared to control (untreated) cells. Each treatment group consists of replicates. Cell growth inhibition is monitored.

In vivo treatment of NIH 3T3 cells overexpressing E-cadherin with anti-E-cadherin monoclonal antibodies.

[00309] NIH 3T3 cells transfected with either a E-cadherin expression plasmid or the neo-DHFR vector are injected into nu/nu (athymic) mice subcutaneously at a dose of 10^6 cells in 0.1 ml of phosphate-buffered saline. On days 0, 1, 5 and every 4 days thereafter, 100 μ g (0.1 ml in PBS) of either an irrelevant or anti-E-cadherin monoclonal antibody of the IG2A subclass is injected intraperitoneally. Tumor occurrence and size are monitored for 1 month.

[00310] All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the above-described modes for carrying out the invention, which are obvious to those skilled in the field of molecular biology or related fields, are intended to be within the scope of the following claims.

CLAIMS

1. A method for diagnosing or detecting in a subject a pancreatic cancer, the method comprising:

determining a test level or test activity of E-cadherin protein in a pancreatic cell from the subject, and

determining a control level or control activity in a pancreatic cell from a healthy subject,

wherein the pancreatic cancer is related to abnormal expression or function of E-cadherin protein, and

wherein the test level or test activity in the cell from the subject is different from the control level or control activity in a pancreatic cell from a healthy subject is indicative of the presence of the pancreatic cancer.

2. The method of claim 1, wherein the level of the E-cadherin protein is determined using an antibody that specifically binds to an antigenic region of E-cadherin.

3. The method according to Claim 1, wherein the E-cadherin protein comprises the amino acid sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2 and SEQ ID NO: 3.

4. The method according to Claim 1, wherein the E-cadherin protein is encoded by a polynucleotide sequence comprising the polynucleotide sequence selected from the group consisting of SEQ ID NO: 4, SEQ ID NO: 5 and SEQ ID NO: 6

5. The method of claim 1, wherein the level of a nucleic acid molecule encoding E-cadherin is determined.

6. The method of claim 5, wherein the level of the nucleic acid molecule is determined by contacting one or more probes that specifically hybridize to the nucleic acid molecule.

7. A method for monitoring treatment of a pancreatic cancer in a subject, wherein the pancreatic cancer is related to abnormal expression or function of E-cadherin protein, the method comprising:

determining a first test level or a first test activity of E-cadherin protein in a pancreatic cell from the subject prior to the treatment,

determining a second test level or a second test activity of E-cadherin protein in a pancreatic cell from the subject subsequent to the treatment, and

determining a control level or control activity in a pancreatic cell from a healthy subject,

wherein the second test level or second test activity in the cell from the subject approaches the control level or control activity when compared to the first test level or first test activity is indicative of successful treatment.

8. A method according to Claim 1, wherein the method determines recurrence of the pancreatic cancer.

9. A pharmaceutical composition comprising an antagonist to E-cadherin and a pharmaceutically acceptable excipient.

10. A pharmaceutical composition according to Claim 9, wherein the antagonist is an anti-E-cadherin antibody.

11. A pharmaceutical composition according to Claim 9, wherein the antagonist is an anti-sense nucleic acid molecule or an RNAi molecule that inhibits the translation or transcription of a gene that codes for the E-cadherin protein.

12. A pharmaceutical composition according to Claim 9, wherein the E-cadherin protein comprises the amino acid sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2 and SEQ ID NO: 3.

13. A pharmaceutical composition according to Claim 9, wherein the E-cadherin protein is encoded by a polynucleotide sequence comprising the polynucleotide sequence selected from the group consisting of SEQ ID NO: 4, SEQ ID NO: 5 and SEQ ID NO: 6.

14. A method for treating a pancreatic cancer, wherein the pancreatic cancer is related to abnormal expression or function of E-cadherin protein in a pancreatic cell, the method comprising administering to a patient in need thereof an effective amount of the pharmaceutical composition according to Claim 9.

15. A method of inhibiting cell growth or proliferation comprising contacting cells with E-cadherin antibody.

16. A method of inhibiting cell growth or proliferation comprising contacting cells with E-cadherin RNAi.

17. A method for screening for an agent that modulates E-cadherin protein activity, the method comprising:

(i) contacting a candidate agent with a preparation of E-cadherin protein,
and

(ii) assaying for a E-cadherin protein activity, wherein a change in said E-cadherin protein activity in the presence of said agent relative to a E-cadherin protein activity in the absence of said agent indicates said agent modulates E-cadherin protein activity.

18. A method for screening for an agent that modulates the level of expression of a nucleic acid that codes for a E-cadherin protein in a cell the naturally expresses the E-cadherin protein, the method comprises:

(i) contacting a candidate agent with the cell or a cell-free preparation from the cell wherein E-cadherin protein is expressed, and

(ii) assaying for the level of expression of the E-cadherin protein activity, wherein a change in said level in the presence of said agent relative to a level in the absence of said agent indicates said agent modulates the expression of E-cadherin protein.

19. A method according to Claim 18, wherein the cell is a pancreatic cell.

1/2

Effect of Antibodies on Cell Invasion

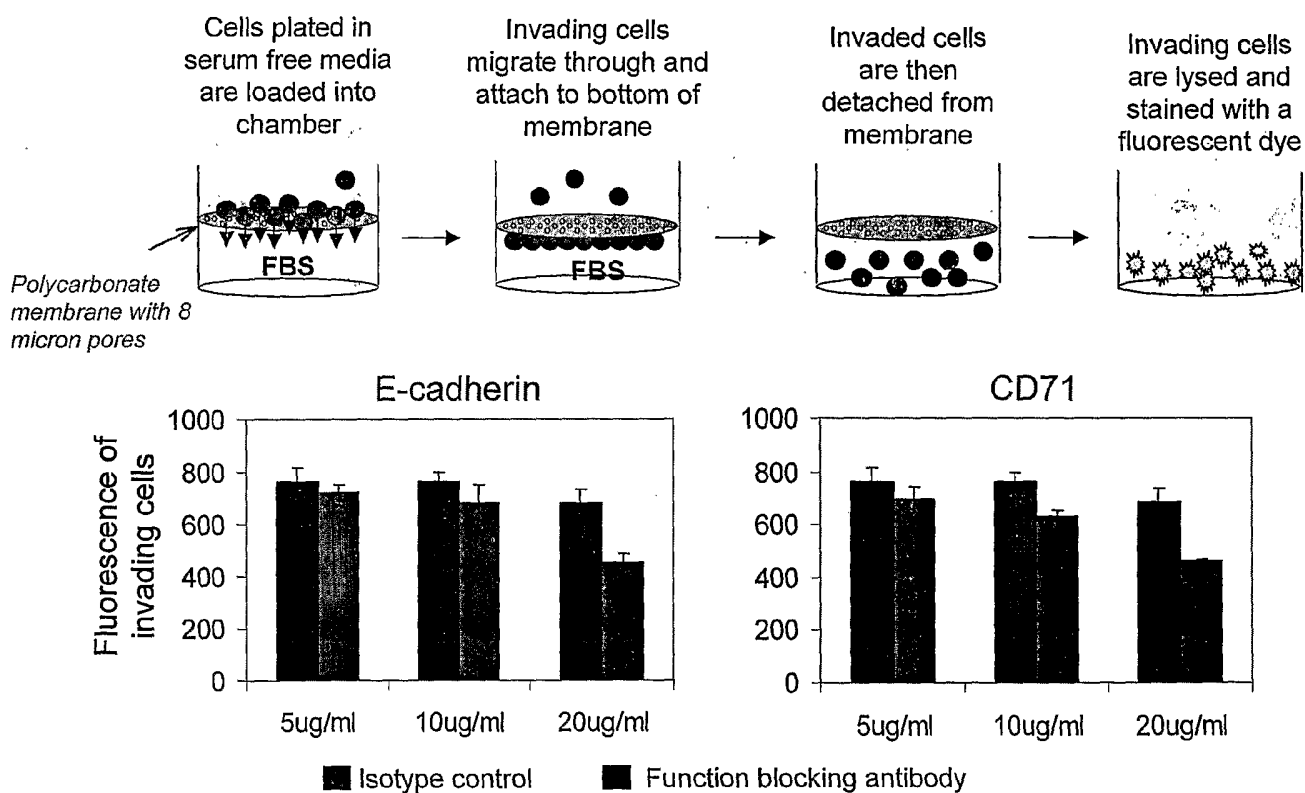


FIGURE 1

2/2

Effect of Antibodies on Apoptosis of Pancreatic Cells

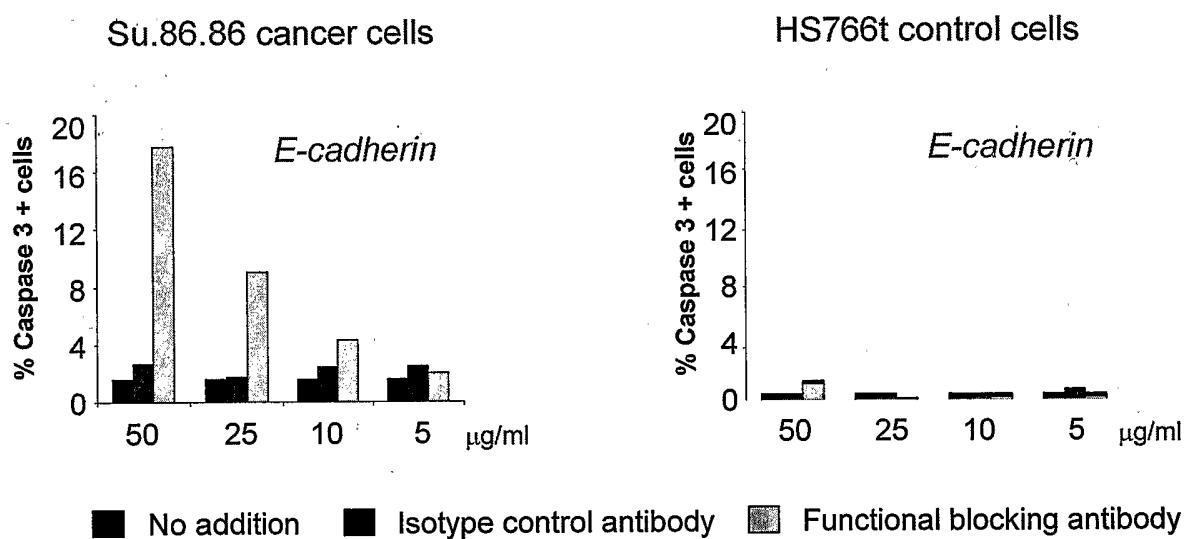


FIGURE 2

SEQUENCE LISTING

<110> DOMON, Bruno

<120> METHOD AND COMPOSITIONS FOR TREATING
PANCREATIC DISEASES TARGETING E-Cadherin

<130> CL001542

<160> 6

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 633

<212> PRT

<213> Homo sapiens

<400> 1

```

Met Glu Ile Leu Ile Thr Val Thr Asp Gln Asn Asp Asn Lys Pro Glu
 1           5           10           15
Phe Thr Gln Glu Val Phe Lys Gly Ser Val Met Glu Gly Thr Ser Val
          20           25           30
Met Glu Val Thr Ala Thr Asp Ala Asp Asp Val Asn Thr Tyr Asn
          35           40           45
Ala Ala Ile Ala Tyr Thr Ile Leu Ser Gln Asp Pro Glu Leu Pro Asp
          50           55           60
Lys Asn Met Phe Thr Ile Asn Arg Asn Thr Gly Val Ile Ser Val Val
          65           70           75           80
Thr Thr Gly Leu Asp Arg Glu Ser Phe Pro Thr Tyr Thr Leu Val Val
          85           90           95
Gln Ala Ala Asp Leu Gln Gly Glu Gly Leu Ser Thr Thr Ala Thr Ala
          100          105          110
Val Ile Thr Val Thr Asp Thr Asn Asp Asn Pro Pro Ile Phe Asn Pro
          115          120          125
Thr Thr Tyr Lys Gly Gln Val Pro Glu Asn Glu Ala Asn Val Val Ile
          130          135          140
Thr Thr Leu Lys Val Thr Asp Ala Asp Ala Pro Asn Thr Pro Ala Trp
          145          150          155          160
Glu Ala Val Tyr Thr Ile Leu Asn Asp Asp Gly Gly Gln Phe Val Val
          165          170          175
Thr Thr Asn Pro Val Asn Asn Asp Gly Ile Leu Lys Thr Ala Lys Gly
          180          185          190
Leu Asp Phe Glu Ala Lys Gln Gln Tyr Ile Leu His Val Ala Val Thr
          195          200          205
Asn Val Val Pro Phe Glu Val Ser Leu Thr Thr Ser Thr Ala Thr Val
          210          215          220
Thr Val Asp Val Leu Asp Val Asn Glu Ala Pro Ile Phe Val Pro Pro
          225          230          235          240
Glu Lys Arg Val Glu Val Ser Glu Asp Phe Gly Val Gly Gln Glu Ile
          245          250          255
Thr Ser Tyr Thr Ala Gln Glu Pro Asp Thr Phe Met Glu Gln Lys Ile
          260          265          270
Thr Tyr Arg Ile Trp Arg Asp Thr Ala Asn Trp Leu Glu Ile Asn Pro
          275          280          285

```

Asp Thr Gly Ala Ile Ser Thr Arg Ala Glu Leu Asp Arg Glu Asp Phe
 290 295 300
 Glu His Val Lys Asn Ser Thr Tyr Thr Ala Leu Ile Ile Ala Thr Asp
 305 310 315 320
 Asn Gly Ser Pro Val Ala Thr Gly Thr Gly Thr Leu Leu Leu Ile Leu
 325 330 335
 Ser Asp Val Asn Asp Asn Ala Pro Ile Pro Glu Pro Arg Thr Ile Phe
 340 345 350
 Phe Cys Glu Arg Asn Pro Lys Pro Gln Val Ile Asn Ile Ile Asp Ala
 355 360 365
 Asp Leu Pro Pro Asn Thr Ser Pro Phe Thr Ala Glu Leu Thr His Gly
 370 375 380
 Ala Ser Ala Asn Trp Thr Ile Gln Tyr Asn Asp Pro Thr Gln Glu Ser
 385 390 395 400
 Ile Ile Leu Lys Pro Lys Met Ala Leu Glu Val Gly Asp Tyr Lys Ile
 405 410 415
 Asn Leu Lys Leu Met Asp Asn Gln Asn Lys Asp Gln Val Thr Thr Leu
 420 425 430
 Glu Val Ser Val Cys Asp Cys Glu Gly Ala Ala Gly Val Cys Arg Lys
 435 440 445
 Ala Gln Pro Val Glu Ala Gly Leu Gln Ile Pro Ala Ile Leu Gly Ile
 450 455 460
 Leu Gly Gly Ile Leu Ala Leu Leu Ile Leu Ile Leu Leu Leu Leu
 465 470 475 480
 Phe Leu Arg Arg Arg Ala Val Val Lys Glu Pro Leu Leu Pro Pro Glu
 485 490 495
 Asp Asp Thr Arg Asp Asn Val Tyr Tyr Asp Glu Glu Gly Gly Gly
 500 505 510
 Glu Glu Asp Gln Asp Phe Asp Leu Ser Gln Leu His Arg Gly Leu Asp
 515 520 525
 Ala Arg Pro Glu Val Thr Arg Asn Asp Val Ala Pro Thr Leu Met Ser
 530 535 540
 Val Pro Arg Tyr Leu Pro Arg Pro Ala Asn Pro Asp Glu Ile Gly Asn
 545 550 555 560
 Phe Ile Asp Glu Asn Leu Lys Ala Ala Asp Thr Asp Pro Thr Ala Pro
 565 570 575
 Pro Tyr Asp Ser Leu Leu Val Phe Asp Tyr Glu Gly Ser Gly Ser Glu
 580 585 590
 Ala Ala Ser Leu Ser Ser Leu Asn Ser Ser Glu Ser Asp Lys Asp Gln
 595 600 605
 Asp Tyr Asp Tyr Leu Asn Glu Trp Gly Asn Arg Phe Lys Lys Leu Ala
 610 615 620
 Asp Met Tyr Gly Gly Gly Glu Asp Asp
 625 630

<210> 2

<211> 882

<212> PRT

<213> Homo sapiens

<400> 2

Met Gly Pro Trp Ser Arg Ser Leu Ser Ala Leu Leu Leu Leu Leu Gln
 1 5 10 15
 Val Ser Ser Trp Leu Cys Gln Glu Pro Glu Pro Cys His Pro Gly Phe
 20 25 30
 Asp Ala Glu Ser Tyr Thr Phe Thr Val Pro Arg Arg His Leu Glu Arg

3

Glu Asp Phe Gly Val Gly Gln Glu Ile Thr Ser Tyr Thr Ala Gln Glu
 500 505 510
 Pro Asp Thr Phe Met Glu Gln Lys Ile Thr Tyr Arg Ile Trp Arg Asp
 515 520 525
 Thr Ala Asn Trp Leu Glu Ile Asn Pro Asp Thr Gly Ala Ile Ser Thr
 530 535 540
 Arg Ala Glu Leu Asp Arg Glu Asp Phe Glu His Val Lys Asn Ser Thr
 545 550 555 560
 Tyr Thr Ala Leu Ile Ile Ala Thr Asp Asn Gly Ser Pro Val Ala Thr
 565 570 575
 Gly Thr Gly Thr Leu Leu Leu Ile Leu Ser Asp Val Asn Asp Asn Ala
 580 585 590
 Pro Ile Pro Glu Pro Arg Thr Ile Phe Phe Cys Glu Arg Asn Pro Lys
 595 600 605
 Pro Gln Val Ile Asn Ile Ile Asp Ala Asp Leu Pro Pro Asn Thr Ser
 610 615 620
 Pro Phe Thr Ala Glu Leu Thr His Gly Ala Ser Ala Asn Trp Thr Ile
 625 630 635 640
 Gln Tyr Asn Asp Pro Thr Gln Glu Ser Ile Ile Leu Lys Pro Lys Met
 645 650 655
 Ala Leu Glu Val Gly Asp Tyr Lys Ile Asn Leu Lys Leu Met Asp Asn
 660 665 670
 Gln Asn Lys Asp Gln Val Thr Thr Leu Glu Val Ser Val Cys Asp Cys
 675 680 685
 Glu Gly Ala Ala Gly Val Cys Arg Lys Ala Gln Pro Val Glu Ala Gly
 690 695 700
 Leu Gln Ile Pro Ala Ile Leu Gly Ile Leu Gly Gly Ile Leu Ala Leu
 705 710 715 720
 Leu Ile Leu Ile Leu Leu Leu Leu Leu Phe Leu Arg Arg Arg Ala Val
 725 730 735
 Val Lys Glu Pro Leu Leu Pro Pro Glu Asp Asp Thr Arg Asp Asn Val
 740 745 750
 Tyr Tyr Tyr Asp Glu Glu Gly Gly Gly Glu Glu Asp Gln Asp Phe Asp
 755 760 765
 Leu Ser Gln Leu His Arg Gly Leu Asp Ala Arg Pro Glu Val Thr Arg
 770 775 780
 Asn Asp Val Ala Pro Thr Leu Met Ser Val Pro Arg Tyr Leu Pro Arg
 785 790 795 800
 Pro Ala Asn Pro Asp Glu Ile Gly Asn Phe Ile Asp Glu Asn Leu Lys
 805 810 815
 Ala Ala Asp Thr Asp Pro Thr Ala Pro Pro Tyr Asp Ser Leu Leu Val
 820 825 830
 Phe Asp Tyr Glu Gly Ser Gly Ser Glu Ala Ala Ser Leu Ser Ser Leu
 835 840 845
 Asn Ser Ser Glu Ser Asp Lys Asp Gln Asp Tyr Asp Tyr Leu Asn Glu
 850 855 860
 Trp Gly Asn Arg Phe Lys Lys Leu Ala Asp Met Tyr Gly Gly Gly Glu
 865 870 875 880
 Asp Asp

<210> 3

<211> 821

<212> PRT

<213> Homo sapiens

<400> 3

```

Met Gly Pro Trp Ser Arg Ser Leu Ser Ala Leu Leu Leu Leu Leu Gln
 1          5          10          15
Val Ser Ser Trp Leu Cys Gln Glu Pro Glu Pro Cys His Pro Gly Phe
 20          25          30
Asp Ala Glu Ser Tyr Thr Phe Thr Val Pro Arg Arg His Leu Glu Arg
 35          40          45
Gly Arg Val Leu Gly Arg Val Asn Phe Glu Asp Cys Thr Gly Arg Gln
 50          55          60
Arg Thr Ala Tyr Phe Ser Leu Asp Thr Arg Phe Lys Val Gly Thr Asp
 65          70          75          80
Gly Val Ile Thr Val Lys Arg Pro Leu Arg Phe His Asn Pro Gln Ile
 85          90          95
His Phe Leu Val Tyr Ala Trp Asp Ser Thr Tyr Arg Lys Phe Ser Thr
100          105          110
Lys Val Thr Leu Asn Thr Val Gly His His His Arg Pro Pro Pro His
115          120          125
Gln Ala Ser Val Ser Gly Ile Gln Ala Glu Leu Leu Thr Phe Pro Asn
130          135          140
Ser Ser Pro Gly Leu Arg Arg Gln Lys Arg Asp Trp Val Ile Pro Pro
145          150          155          160
Ile Ser Cys Pro Glu Asn Glu Lys Gly Pro Phe Pro Lys Asn Leu Val
165          170          175
Gln Ile Lys Ser Asn Lys Asp Lys Glu Gly Lys Val Phe Tyr Ser Ile
180          185          190
Thr Gly Gln Gly Ala Asp Thr Pro Pro Val Gly Val Phe Ile Ile Glu
195          200          205
Arg Glu Thr Gly Trp Leu Lys Val Thr Glu Pro Leu Asp Arg Glu Arg
210          215          220
Ile Ala Thr Tyr Thr Leu Phe Ser His Ala Val Ser Ser Asn Gly Asn
225          230          235          240
Ala Val Glu Asp Pro Met Glu Ile Leu Ile Thr Val Thr Asp Gln Asn
245          250          255
Asp Asn Lys Pro Glu Phe Thr Gln Glu Val Phe Lys Gly Ser Val Met
260          265          270
Glu Gly Ala Leu Pro Gly Thr Ser Val Met Glu Val Thr Ala Thr Asp
275          280          285
Ala Asp Asp Asp Val Asn Thr Tyr Asn Ala Ala Ile Ala Tyr Thr Ile
290          295          300
Leu Ser Gln Asp Pro Glu Leu Pro Asp Lys Asn Met Phe Thr Ile Asn
305          310          315          320
Arg Asn Thr Gly Val Ile Ser Val Val Thr Thr Gly Leu Asp Arg Glu
325          330          335
Ser Phe Pro Thr Tyr Thr Leu Val Val Gln Ala Ala Asp Leu Gln Gly
340          345          350
Glu Gly Leu Ser Thr Thr Ala Thr Ala Val Ile Thr Val Thr Asp Thr
355          360          365
Asn Asp Asn Pro Pro Ile Phe Asn Pro Thr Thr Gly Leu Asp Phe Glu
370          375          380
Ala Lys Gln Gln Tyr Ile Leu His Val Ala Val Thr Asn Val Val Pro
385          390          395          400
Phe Glu Val Ser Leu Thr Thr Ser Thr Ala Thr Val Thr Val Asp Val
405          410          415
Leu Asp Val Asn Glu Ala Pro Ile Phe Val Pro Pro Glu Lys Arg Val
420          425          430
Glu Val Ser Glu Asp Phe Gly Val Gly Gln Glu Ile Thr Ser Tyr Thr
435          440          445

```

Ala Gln Glu Pro Asp Thr Phe Met Glu Gln Lys Ile Thr Tyr Arg Ile
 450 455 460
 Trp Arg Asp Thr Ala Asn Trp Leu Glu Ile Asn Pro Asp Thr Gly Ala
 465 470 475 480
 Ile Ser Thr Arg Ala Glu Leu Asp Arg Glu Asp Phe Glu His Val Lys
 485 490 495
 Asn Ser Thr Tyr Thr Ala Leu Ile Ile Ala Thr Asp Asn Gly Ser Pro
 500 505 510
 Val Ala Thr Gly Thr Gly Thr Leu Leu Leu Ile Leu Ser Asp Val Asn
 515 520 525
 Asp Asn Ala Pro Ile Pro Glu Pro Arg Thr Ile Phe Phe Cys Glu Arg
 530 535 540
 Asn Pro Lys Pro Gln Val Ile Asn Ile Ile Asp Ala Asp Leu Pro Pro
 545 550 555 560
 Asn Thr Ser Pro Phe Thr Ala Glu Leu Thr His Gly Ala Ser Ala Asn
 565 570 575
 Trp Thr Ile Gln Tyr Asn Asp Pro Thr Gln Glu Ser Ile Ile Leu Lys
 580 585 590
 Pro Lys Met Ala Leu Glu Val Gly Asp Tyr Lys Ile Asn Leu Lys Leu
 595 600 605
 Met Asp Asn Gln Asn Lys Asp Gln Val Thr Thr Leu Glu Val Ser Val
 610 615 620
 Cys Asp Cys Glu Gly Ala Ala Gly Val Cys Arg Lys Ala Gln Pro Val
 625 630 635 640
 Glu Ala Gly Leu Gln Ile Pro Ala Ile Leu Gly Ile Leu Gly Gly Ile
 645 650 655
 Leu Ala Leu Leu Ile Leu Ile Leu Leu Leu Leu Phe Leu Arg Arg
 660 665 670
 Arg Ala Val Val Lys Glu Pro Leu Leu Pro Pro Glu Asp Asp Thr Arg
 675 680 685
 Asp Asn Val Tyr Tyr Tyr Asp Glu Glu Gly Gly Gly Glu Glu Asp Gln
 690 695 700
 Asp Phe Asp Leu Ser Gln Leu His Arg Gly Leu Asp Ala Arg Pro Glu
 705 710 715 720
 Val Thr Arg Asn Asp Val Ala Pro Thr Leu Met Ser Val Pro Arg Tyr
 725 730 735
 Leu Pro Arg Pro Ala Asn Pro Asp Glu Ile Gly Asn Phe Ile Asp Glu
 740 745 750
 Asn Leu Lys Ala Ala Asp Thr Asp Pro Thr Ala Pro Pro Tyr Asp Ser
 755 760 765
 Leu Leu Val Phe Asp Tyr Glu Gly Ser Gly Ser Glu Ala Ala Ser Leu
 770 775 780
 Ser Ser Leu Asn Ser Ser Glu Ser Asp Lys Asp Gln Asp Tyr Asp Tyr
 785 790 795 800
 Leu Asn Glu Trp Gly Asn Arg Phe Lys Lys Leu Ala Asp Met Tyr Gly
 805 810 815
 Gly Gly Glu Asp Asp
 820

<210> 4

<211> 4826

<212> DNA

<213> Homo sapiens

<400> 4

agtggcgtcg gaactgcaaa gcacctgtga gcttgcgga gtcagttcag actccagccc 60

```

gctccagccc ggcccgaccc gaccgcaccc ggcgctgccc ctcgctcggc gtccccggcc 120
agccatgggg ccttgagacc gcagcctctc ggcgctgctg ctgctgctgc aggtctcctc 180
ttggctctgc caggagccgg agccctgcca ccctggcttt gacgccgaga gctacacgtt 240
cacggtgccc cggcgccacc tggagagagg ccgctcctg ggcagagctc tgtgatttct 300
gccctgcagt gaattttgaa gattgcaccg gtcgacaaag gacagcctat ttttccctcg 360
acacccgatt caaagtgggc acagatggtg tgattacagt caaaaggcct ctacggtttc 420
ataaccacac gatccatttc ttggtctacg cctgggactc cactacaga aagttttcca 480
ccaaagtcac gctgaataca gtggggcacc accaccgcc cccgccccat caggcctccg 540
tttctggaat ccaagcagaa ttgctcacat ttcccaactc ctctcctggc ctcagaagac 600
agaagagaga ctgggttatt cctcccatca gctgcccaga aaatgaaaaa ggcccatbbc 660
ctaaaaacct gggtcagatc aaatccaaca aagacaaaga aggcaagggt ttctacagca 720
tactggcca aggagctgac acacccctg ttggtgtctt tattattgaa agagaaacag 780
gatggctgaa ggtgacagag cctctggata gagaacgcat tgccacatac actctcttct 840
ctcacgctgt gtcacccaac gggaatgcag ttgaggatcc aatggagatt ttgatcacgg 900
taaccgatca gaatgacaac aagcccgaat tcaccagga ggtctttaag gggctctgtc 960
tggaaggaac ctctgtgatg gaggtcacag ccacagacgc ggacgatgat gtgaacacct 1020
acaatgccgc catcgcttac accatcctca gccaaatcc tgagctccct gacaaaaata 1080
tggtcaccat taacaggaac acaggagtca tcagtgtggt caccactggg ctggaccgag 1140
agagtttccc tacgtatacc ctggtggttc aagctgctga ccttcaagggt gagggttaa 1200
gcacaacagc aacagctgtg atcacagtca ctgacaccaa cgataatcct ccgatcttca 1260
atcccaccac gtacaagggt caggtgctg agaacgaggc taacgtcgta atcaccacac 1320
tgaaagtgac tgatgctgat gcccacaata cccagcgtg ggaggctgta tacaccatat 1380
tgaatgatga tgggtggaca tttgtcgtca ccacaaatcc agtgaacaac gatggcattt 1440
tgaaaacagc aaagggttgg gattttgagg ccaagcagca gtacattcta cacgtagcag 1500
tgacgaatgt ggtacctttt gaggtctctc tcaccacctc cacagccacc gtcaccgtgg 1560
atgtgctgga tgtgaatgaa gcccctatct ttgtgcctcc tgaaaagaga gtggaagtgt 1620
ccgaggactt tggcgtgggc caggaaatca catcctacac tgcccaggag ccagacacat 1680
ttatggaaca gaaaataaca tatcgattt ggagagacac tgccaactgg ctggagatta 1740
atccggacac tgggtccatt tccactcggg ctgagctgga cagggaggat tttgagcacg 1800
tgaagaacag caggtacaca gccctaatac tagctacaga caatggttct ccagttgcta 1860
ctggaacagg gacacttctg ctgatcctgt ctgatgtgaa tgacaacgcc cccataccag 1920
aacctcgaac tatattcttc tgtgagagga atccaaagcc tcaggtcata aacatcattg 1980
atgcagacct tcctcccaat acatctccct tcacagcaga actaacacac ggggcgagtg 2040
ccaactggac cattcagtac aacgaccaa cccaagaatc tatcattttg aagccaaaga 2100
tggccttaga ggtgggtgac taaaaaatca atctcaagct catggataac cagaataaag 2160
accaagtgac caccttagag gtcagcgtgt gtgactgtga aggggccgct ggcgtctgta 2220
ggaaggcaca gcctgtcgaa gcaggattgc aaattcctgc cattctgggg attcttggag 2280
gaattcttgc tttgctaatt ctgattctgc tgccttctgt gtttcttcgg aggagagcgg 2340
tggtcaaaga gcccttactg ccccagagg atgacacccg ggacaacgtt tattactatg 2400
atgaagaagg aggcggagaa gaggaccagg actttgactt gagccagctg cacagggggc 2460
tggaagcctg gcctgaagtg actcgtaacg acgttgccac aacctcatg agtgcccc 2520
ggtatcttcc ccgccctgcc aatcccgatg aaattggaaa ttttattgat gaaaatctga 2580
aagcggctga tactgacccc acagccccgc cttatgattc tctgctcgtg tttgactatg 2640
aaggaagcgg ttccgaagct gctagtctga gctccctgaa ctctcagag tcagacaaag 2700
accaggacta tgactacttg aacgaatggg gcaatcgctt caagaagctg gctgacatgt 2760
acggaggcgg cgaggacgac taggggactc gagagaggcg ggcccagac ccatgtgctg 2820
ggaaatgcag aaatcacgtt gctggtggtt ttccagctcc cttcccttga gatgagtttc 2880
tggggaaaaa aaagagactg gttagtgatg cagttagtat agctttatac tctctccact 2940
ttatagctct aataagtttg tggtagaaaa gtttcgactt atttcttaaa gctttttttt 3000
ttttccatc actctttaca tgggtggtgat gtccaaaaga tacccaaatt ttaatatcc 3060
agaagaacaa ctttagcatc agaaggttca cccagcacct tgcaattttt cttaaggaat 3120
tttgtctcac ttttaaaaag aaggggagaa gtcagctact ctagtctctg tgttttgtgt 3180
atataatttt ttaaaaaaaa tttgtgtgct tctgctcatt actacactgg tgtgtccctc 3240
tgcctttttt ttttttttaa gacagggtct cattctatcg gccaggctgg agtgagtggt 3300
tgcaatcaca gctcactgca gccttgtcct cccaggctca agctatcctt gcacctcagc 3360
ctcccaagta gctgggacca caggcatgca ccaactacga tgactaattt ttaaatatt 3420
tgagacgggg tctccctgtg ttaccaggc ttggtctcaa ctcctgggct caagtgatcc 3480

```

```

tcccattcttg gcctcccaga gtattgggat tacagacatg agccactgca cctgcccagc 3540
tccccaactc cctgccattt ttttaagagac agtttcgctc catcgcccag gcctgggatg 3600
cagtgatgtg atcatagctc actgtaacct caaactcttg ggctcaagca gttctccac 3660
cagcctcctt tttatttttt tgtacagatg gggctcttgc atgttgccca agctggctct 3720
aaactcctgg cctcaagcaa tccttctgcc ttggccccc aaagtgtctg gattgtgggc 3780
atgagctgct gtgccagacc tccatgtttt aatatcaact ctcaactctg aattcagttg 3840
ctttgcccga gataggagtt ctctgatgca gaaattattg ggctctttta gggtaagaag 3900
tttgtgtctt tgtctggcca catcttgact aggtattgtc tactctgaag acctttaatg 3960
gcttccctct ttcattctct gagtatgtaa cttgcaatgg gcagctatcc agtgacttgt 4020
tctgagtaag tgtgttcatt aatgtttatt tagctctgaa gcaagagtga tatactccag 4080
gacttagaat agtgcctaaa gtgctgcagc caaagacaga gcggaactat gaaaagtggg 4140
cttgagagat gcaggagagc ttgtcattga gcctggcaat ttagcaaact gatgctgagg 4200
atgattgagg tgggtctacc tcattctctga aaattctgga aggaatggag gagtctcaac 4260
atgtgtttct gacacaagat ccgtggtttg tactcaaagc ccagaatccc caagtgcctg 4320
cttttgatga tgtctacaga aaatgctggc tgagctgaac acatttgccc aattccaggt 4380
gtgcacagaa aaccgagaat attcaaaatt ccaaattttt ttcttaggag caagaagaaa 4440
atgtggccct aaagggggtt agttgagggg taggggtag tgaggatctt gatttggatc 4500
tctttttatt taaatgtgaa tttcaacttt tgacaatcaa agaaaagact tttgtgaaa 4560
tagctttact gtttctcaag tgttttgag aaaaaaatca accctgcaat cacttttttg 4620
aattgtcttg atttttcggc agttcaagct atatcgaata tagttctgtg tagagaatgt 4680
cactgtagtt ttgagtgtat acatgtgtgg gtgctgataa ttgtgtattt tctttggggg 4740
tggaagagga aaacaattca agctgagaaa agtattctca aagatgcatt tttataaatt 4800
ttattaaaca attttgtaa accatt 4826

```

<210> 5

<211> 4816

<212> DNA

<213> Homo sapiens

<400> 5

```

agtggcgctg gaactgcaaa gcacctgtga gcttgcgga gtcagttcag actccagccc 60
gctccagccc ggcccgaccc gaccgcaccc ggcgcctgcc ctgctcgcc gtccccggcc 120
agccatgggc ccttgagacc gcagcctctc ggcgctgctg ctgctgctgc aggtctctct 180
ttggctctgc caggagccgg agccctgcca ccctggcttt gacgccgaga gctacacggt 240
cacggtgccc cggcgccacc tggagagagg ccgctgctct ggcagagtga attttgaaga 300
ttgcaccggt cgacaaagga cagcctatct ttccctcgac acccgattca aagtgggcac 360
agatggtgtg attacagtca aaaggcctct acggtttcat aaccacaga tccatttctt 420
ggtctacgcc tgggactcca cctacagaaa gttttccacc aaagtcacgc tgaatacagt 480
ggggcaccac caccgcccc cgccccatca ggcctccgtt tctggaatcc aagcagaatt 540
gtccacattt cccaactcct ctccctggcct cagaagacag aagagagact gggttatttc 600
tcccatacgc tgcccagaaa atgaaaaagg cccatttctt aaaaacctgg ttcagatcaa 660
atccaacaaa gacaaagaag gcaaggtttt ctacagcatc actggccaag gagctgacac 720
accccctgtt ggtgtcttta ttattgaaag agaaacagga tggctgaagg tgacagagcc 780
tctggataga gaacgcattg ccacatacac tctcttctct cacgctgtgt catccaacgg 840
gaatgcagtt gaggatccaa tggagatttt gatcacggta accgatcaga atgacaacaa 900
gcccgaattc acccaggagg tctttaaggg gtctgtcatg gaagggtgctc ttccagggaac 960
ctctgtgatg gaggtcacag ccacagacgc ggacgatgat gtgaacacct acaatgccgc 1020
catcgcttac accatcctca gccaaagatc tgagctccct gacaaaaata tggtcaccat 1080
taacaggaac acaggagtca tcagtgtggt caccactggg ctggaccgag agagtctccc 1140
tacgtatacc ctggtggttc aagctgctga ccttcaaggt gaggggttaa gcacaacagc 1200
aacagctgtg atcacagtca ctgacaccaa cgataatcct ccgatcttca atcccaccac 1260
gtacaagggt caggtgcctg agaacgaggc taacgtcgta atcaccacac tgaaagtgc 1320
tgatgctgat gcccacaata cccagcgtg ggaggctgta tacaccatat tgaatgatga 1380
tgggtggaca tttgtcgtca ccacaaatcc agtgaacaac gatggcattt tgaaaacagc 1440
aaagggcttg gattttgagg ccaagcagca gtacattcta cacgtagcag tgacgaatgt 1500
ggtacctttt gaggtctctc tcaccacctc cacagccacc gtcaccgtgg atgtgctgga 1560
tgtgaatgaa gcccctatct ttgtgcctcc tgaaaagaga gtggaagtgt ccgaggactt 1620

```

```

tggcgtgggc caggaaatca catcctacac tgcccaggag ccagacacat ttatggaaca 1680
gaaaataaca tatcggatth ggagagacac tgccaactgg ctggagatta atccggacac 1740
tggtgccatt tccactcggg ctgagctgga cagggaggat tttgagcacg tgaagaacag 1800
cacgtacaca gccctaataca tagctacaga caatggttct ccagttgcta ctggaacagg 1860
gacacttctg ctgatcctgt ctgatgtgaa tgacaacgcc cccataccag aacctcgaac 1920
tatattcttc tgtgagagga atccaaagcc tcaggtcata aacatcattg atgcagacct 1980
tcctcccaat acatctccct tcacagcaga actaacacac ggggagagtg ccaactggac 2040
cattcagtac aacgacccaa cccaagaatc tatcattttg aagccaaaga tggccttaga 2100
ggtgggtgac tacaaaaatca atctcaagct catggataac cagaataaag accaagtgc 2160
caccttagag gtcagcgtgt gtgactgtga agggccgct ggcgtctgta tgaagtcaca 2220
gcctgtcgaa gcaggattgc aaattcctgc cattctgggg attcttgagg gaattcttgc 2280
tttgctaatt ctgattctgc tgccttctgc gtttcttcgg aggagagcgg tggtaaaaga 2340
gcccttactg ccccagagg atgacacccg ggacaacgtt tattactatg atgaagaagg 2400
aggcggagaa gaggaccagg actttgactt gagccagctg cacagggggc tggacgctcg 2460
gcctgaagtg actcgtaacg acgttgccac aacctctatg agtgcccc ggtatcttcc 2520
ccgcctgcc aatcccgatg aaattggaaa ttttattgat gaaaatctga aagcggctga 2580
tactgacccc acagccccgc cttatgattc tctgctcgtg tttgactatg aaggaaagg 2640
ttccgaagct gctagtctga gctccctgaa ctctcagag tcagacaaag accaggacta 2700
tgactacttg aacgaatggg gcaatcgctt caagaagctg gctgacatgt acggaggcgg 2760
cgaggacgac taggggactc gagagaggcg ggcaccagac ccagtgtctg ggaatgcag 2820
aaatcacgtt gctgggtggt tttcagctcc ctcccttga gatgagttt tggggaaaaa 2880
aaagagactg gttagtgatg cagttagtat agctttatac tctctccact ttatagctct 2940
aataagtttg tgttagaaaa gtttcgactt atttcttaaa gctttttttt ttttccatc 3000
actctttaca tgggtggtgat gtccaaaaga taccacaaat ttaatatctc agaagaacaa 3060
ctttagcatc agaaggttca cccagcacct tgcagatttt cttaagggaat tttgtctcac 3120
ttttaaaaag aaggggagaa gtcagctact ctagtctgtg tgttttgtgt atataatttt 3180
ttaaaaaaaa tttgtgtgct tctgctcatt actacactgg tgtgtccctc tgcctttttt 3240
ttttttttta gacagggtct cattctatcg gccaggctgg agtgagtggt tgcaatcaca 3300
gctcactgca gccttgcct cccaggetca agctatcctt gcacctcagc ctcccaagta 3360
gctgggacca caggcatgca ccactacgca tgactaatth tttaaatatt tgagacgggg 3420
tctccctgtg ttaccagggc tgggtctcaaa ctccctgggt caagtgatcc tcccatcttg 3480
gcctcccgaa gtattgggat tacagacatg agccactgca cctgcccagc tccccaactc 3540
cctgccattt tttaagagac agtttcgctc catcgcccag gcctgggatg cagtgatgtg 3600
atcatagctc actgtaacct caaactctgg ggtcaagca gttctccac cagcctcctt 3660
tttatttttt tgtacagatg gggctctgct atgttgcca agctgggtct aaactcctgg 3720
cctcaagcaa tccttctgcc ttggccccc ctaactcctg aattcagttg ctttgcccaa 3840
gtgcccagcc tccatgtttt gaaattattg ggctctttta gggtaagaag tttgtgtctt 3900
gataggagtt ctctgatgca aggtattgtc tactctgaag acctttaatg gcttccctct 3960
tgtctggcca catcttgact aggtattgtc ctgcaatgg gcagctatcc agtgacttgt tctgagtaag 4020
tgtgttcatt aatgtttatt tagctctgaa gcaagagtga tatactccag gacttagaat 4080
agtgccataa gtgctgcagc caaagacaga gcggaactat gaaaagtggg cttggagatg 4140
gcaggagagc ttgtcattga gcctggcaat ttagcaaaact gatgctgagg atgattgagg 4200
tgggtctacc tcatctctga aaattctgga agaatggag gagtctcaac atgtgtttct 4260
gacacaagat ccgtggtttg tactcaaagc ccagaatccc caagtgcctg cttttgatga 4320
tgtctacaga aaatgtggtg tgagctgaac acatttgccc aattccaggt gtgcacagaa 4380
aaccgagaat attcaaaatt ccaaattttt ttcttaggag caagaagaaa atgtggcct 4440
aaaggggggt agttgagggg tagggggtag tgaggatctt gatttgatc tctttttatt 4500
tāaatgtgaa tttcaacttt tgacaatcaa agaaaagact tttgttgaaa tagctttact 4560
gtttctcaag tgttttgagg aaaaaatca accctgcaat cactttttgg aattgtcttg 4620
atttttcggc agttcaagct atatcgata tagttctgtg tagagaatgt cactgagtt 4680
ttgagtgat acatgtgtgg gtgctgataa ttgtgtatt tctttggggg tggaaaagga 4740
aaacaattca agctgagaaa agtattctca aagatgcatt tttataaatt ttattaaaca 4800
attttgtaa accatt 4816

```

<210> 6

<211> 4633

<212> DNA

<213> Homo sapiens

<400> 6

```

agtggcgtcg gaactgcaaa gcacctgtga gcttgccgaa gtcagttcag actccagccc 60
gctccagccc ggcccgaccc gaccgcaccc ggccgctgcc ctgcgtcggc gtccccggcc 120
agccatggggc ccttggagcc gcagcctctc ggccgctgctg ctgctgctgc aggtctcctc 180
ttggctctgc caggagccgg agccctgcc ccttggcttt gacgccgaga gctacacgtt 240
cacggtgccc cggcgccacc tggagagagg ccgcgtcctg ggcagagtga attttgaaga 300
ttgcaccggt cgacaaagga cagcctatct tccctcgac acccgattca aagtgggcac 360
agatggtgtg attacagtca aaaggcctct acggtttcat aaccacaga tccatttctt 420
ggtctacgcc tgggactcca cctacagaaa gttttccacc aaagtcacgc tgaatacagt 480
ggggcaccac caccgcccc cgcccatca ggccctcgtt tctggaatcc aagcagaatt 540
gctcacatct cccaactcct ctctggcctc cagaagacag aagagagact ggtttattcc 600
tcccacagc tgcacagaaa atgaaaaagg cctatttctt aaaaacctgg ttcagatcaa 660
atccaacaaa gacaaagaag gcaaggtttt ctacagcatc actggccaag gactgacac 720
acccctgtt ggtgtcttta ttattgaaag agaaacagga tggctgaagg tgacagagcc 780
tctgtagata gaacgcattg ccacatacac tctcttctct cagctgtgtg catccaacgg 840
gaatgcagtt gaggatccaa tggagatctt gatcacggta accgatcaga atgacaacaa 900
gcccgaattc acccaggagg tctttaaggg gtctgtcatg gaagtgctc tccaggaac 960
ctctgtgatg gaggtcacag ccacagacgc ggacgatgat gtgaacacct acaatgccgc 1020
catcgcttac accatcctca gccaatgccc gacaaaaata tgttcacat 1080
taacaggaac acaggagtca tcaagtgtgt caccactggg ctggaccgag agagtttccc 1140
tacgtatacc ctggtggttc aagctgctga ccttcaaggg gaggggttaa gcacaacagc 1200
aacagctgtg atcacagtca ctgacaccaa cgataatcct ccgatcttca atcccaccac 1260
gggcttggtt tttgaggcca agcagcagta cattctacac gtagcagtga cgaatgtggt 1320
accttttgag gtctctctca ccacctccac agccaccgct accgtggatg tgctggatgt 1380
gaatgaagcc cccatctttg tgcctcctga aaagagagtg gaagtgtccg aggacttttg 1440
cgtgggccag gaaatcacat cctacactgc ccaggagcca gacacattta tggacagaa 1500
aataacatat cggatttggg gagacactgc caactggctg gagattaatc cggacactgg 1560
tgccatttcc actcgggctg agctggacag ggaggatttt gagcacgtga agaacagcac 1620
gtacacagcc taatcatag ctacagacaa tggttctcca gttgctactg gaacagggac 1680
acttctgtg atcctgtctg atgtgaatga caacgcccc ataccagaac ctggaactat 1740
atttctctgt gagaggaatc caaagcctca ggtcataaac atcattgatg cagaccttcc 1800
tcccataaca tctcccttca cagcagaact aacacacggg gcgagtacca actggaccat 1860
tcagtacaac gacccaaccc aagaatctat ctttttgaag ccaaagatgg ccttagaggt 1920
gggtgactac aaaatcaatc tcaagctcat ggataaccag aataaagacc aagtgaccac 1980
cttagaggtc agcgtgtgtg actgtgaagg ggccgctggc gtctgttagga aggcacagcc 2040
tgtcgaagca ggattgcaaa ttcttgccat tctggggatt cttggaggaa ttcttgcttt 2100
gctaattctg attctgctgc tcttgctgtt tcttcggagg agagcgggtg tcaaagagcc 2160
cttactgccc ccagaggatg acacccggga caacgtttat tactatgatg aagaaggagg 2220
cggagaagag gaccaggact ttgacttgag ccagctgcac aggggcctgg acgctcggcc 2280
tgaagtgact cgtaacgacg ttgcaccaac cctcatgagt gtcccccggt atcttccccg 2340
ccctgccaat cccgatgaaa ttggaatttt tattgatgaa aatctgaaag cggctgatac 2400
tgacccaca gccccgcctt atgattctct gctcgtgttt gactatgaag gaagcggttc 2460
cgaagctgct agtctgagct cctgaactc ctacagatca gacaaagacc aggactatga 2520
ctacttgaac gaatggggca atcgcttcaa gaagctggct gacatgtacg gaggcggcga 2580
ggacgactag gggactcgag agaggcgggc ccagaccca tgtgctggga aatgcagaaa 2640
tcacgttgct ggtggttttt cagctccctt cccttgagat gattttctgg ggaaaaaaaa 2700
gagactggtt agtgatgcag ttagtatagc tttatactct ctccacttta tagctcta 2760
aagtttgtgt tagaaaagt tgcacttatt tcttaaagct tttttttttt tcccatcact 2820
ctttacatgg tggatgctc caaaagatac ccaaatttta atattccaga agaacaactt 2880
tagcatcaga aggttcaccc agcaccttgc agattttctt aagggaatttt gctcactttt 2940
taaaaagaag gggagaagtc agctactcta gttctgttgt tttgtgtata taatttttta 3000
aaaaaaaaat ttgtgcttct gctcattact acactgggtg gtccctctgc cttttttttt 3060
tttttaagac agggctctcat tctatcgccc aggttgaggt gcagtgggtg aatcacagct 3120
cactgcagcc ttgtcctccc aggtcgaagc tatccttgca cctcagcctc ccaagtagct 3180

```

gggaccacag gcatgcacca ctacgcatga ctaatTTTTT aaatatTTTga gacggggTct 3240
ccctgtgtta cccaggctgg tctcaaactc ctgggctcaa gtgatccctc catcttggcc 3300
tcccagagta ttgggattac agacatgagc cactgcacct gccagctcc ccaactccct 3360
gccatTTTTt aagagacagt ttcgctccat cgcccaggcc tgggatgcag tgatgtgatc 3420
atagctcact gtaacctcaa actctggggc tcaagcagtt ctcccaccag cctcctTTTT 3480
atTTTTTgt acagatgggg tcttgctatg ttgcccaagc tggTcttaaa ctcctggcct 3540
caagcaatcc ttctgccttg gcccccaaa gtgctgggat tgtgggcatg agctgctgtg 3600
cccagcctcc atgtTTTaat atcaactctc actcctgaat tcagttgctt tgcccaagat 3660
aggagtTctc tgatgcagaa attattgggc tctTTtaggg taagaagttt gtgtctttgt 3720
ctggccacat cttgactagg tattgtctac tctgaagacc tttaatggct tccctcttTc 3780
atctcctgag tatgtaaactt gcaatgggca gctatccagt gacttgtTct gagtaagtgt 3840
gttcattaat gtttattttag ctctgaagca agagtgtat actccaggac ttagaatagt 3900
gcctaaagtg ctgcagccaa agacagagcg gaactatgaa aagtgggctt ggagatggca 3960
ggagagcttg tcattgagcc tggcaattta gcaactgat gctgaggatg attgaggtgg 4020
gtctacctca tctctgaaaa ttctggaagg aatggaggag tctcaacatg tgtttctgac 4080
acaagatccg tggtttTgtac tcaaagccca gaatcccaa gtgcctgctt ttgatgatgt 4140
ctacagaaaa tgctggctga gctgaacaca ttTgccaat tccaggtgtg cacagaaaac 4200
cgagaatatt caaaattcca aatTTTTTt ttaggagcaa gaagaaaatg tggccctaaa 4260
gggggttagt tgaggggtag ggggtagtga ggatcttgat ttggatctct tttattTaa 4320
atgtgaattt caactTTTga caatcaaaga aaagactTTt gttgaaatag ctttactgtt 4380
tctcaagtgt tttggagaaa aaaatcaacc ctgcaatcac tttttggaat tgtcttgatt 4440
tttcggcagt tcaagctata tcgaatatag ttctgtgtag agaatgtcac tgtagttttg 4500
agtgtataca tgtgtgggtg ctgataattg tgtattTtct ttgggggtgg aaaaggaaaa 4560
caattcaagc tgagaaaagt attctcaaag atgcattTTt ataaattTta ttaaacaatt 4620
ttgttaaacc att 4633

