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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 February 2004 (12.02.2004)

PCT

(10) International Publication Number
WO 2004/012574 A2

(51) International Patent Classification[7]: A61B

(21) International Application Number: PCT/US2003/022241

(22) International Filing Date: 16 July 2003 (16.07.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/396,171 16 July 2002 (16.07.2002) US

(71) Applicant: **DELTAGEN PROTEOMICS INC.** [US/US]; 615 Arapeen Drive, Suite 300, Salt Lake City, UT 84108 (US).

(72) Inventor: **KAMB, Carl, Alexander**; 696 Donner Hill Circle, Salt Lake City, UT 84108 (US).

(74) Agent: **WILLIAMS, Joseph, A.**; Marshall, Gerstein & Borun LLP, 6300 Sears Tower, 233 South Wacker Drive, Chicago, Il 60606 (US).

(81) **Designated States** *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, 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, 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, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (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:** NEGATIVE SELECTIONS ASSAYS, AND COMPOSITIONS THEREOF

(57) **Abstract:** The present invention is directed to methods for performing negative selection assays leading to the identification of cytostatic or cytotoxic agents that cause a lethal phenotype. The invention is useful also for evaluation of conditional cytotoxicity and cell-specific cytotoxicity.

WO 2004/012574 A2

NEGATIVE SELECTIONS ASSAYS, AND COMPOSITIONS THEREOF

This application claims priority from U.S. provisional patent application Serial No. 60/396,171 filed July 16, 2002, the entire disclosure of which is specifically incorporated by reference herein in its entirety.

5

BACKGROUND OF THE INVENTION

Colorectal cancer is the second leading cause of cancer-related deaths in the United States, being responsible for as many as 60,000 fatalities each year. Nearly five percent of the US population develops colorectal cancer, and this number is predicted to rise as the average life expectancy increases (Beart, R.W. (1991) *American Cancer Society*
10 *Textbook of Clinical Oncology. Atlanta,* American Cancer Society, pg. 213-218).

Colon cancer erupts from the lumenal surface of the colon and rectum. Normally, epithelial cells line the surface of the colon and invaginate into structures called crypts. Over the course of three to six days, stem cells located at the base of each crypt divide, and then differentiate as they migrate toward the apex where they die and are released
15 into the lumen (see, for example, Lipkin, M. et al. (1963) "Cell proliferation kinetics in the gastrointestinal tract of man." *J Clin Invest* 42:767). The first manifestations of colorectal cancer are often observed clinically as a polyp; a mass of epithelial cells that protrude from the apex of the colonic crypts of the bowel wall (see, for example, Kent, T.H. et al. (1983) "Polyps of the colon and small bowel, polyp syndromes, and the polyp carcinoma sequence."
20 in Norris HT (eds) *Pathology of the Colon, Small Intestine, and Anus.* New York, Churchill Livingstone, vol 2, pg 167). Polyps are, predominantly, divided into two classes. The nondysplastic form consists of a large mass of cells that have normal morphology. These aggregates line up in a single row along the basement membrane and exhibit a low frequency of becoming neoplastic. The second form of polyp is the adenomatous polyp. These
25 formations are dysplastic in nature and exhibit an abnormal intracellular organization. As tumor progression evolves, adenomatous polyps exhibit a high frequency of metastasis to surrounding tissues with the most common sites of invasion being the mesenteric lymph nodes, the peritoneal surface, and the liver.

Genetic studies have shown that at least two forms of heritable colorectal
30 cancers exist. The first group, which includes familial adenomatous polyposis coli (FAP), Peutz-Jegher syndrome, familial juvenile polyposis, Cronkhite-Canada syndrome and hyperplastic polyposis, is characterized by the appearance of multiple (hundreds to

thousands) of benign, precursor, colorectal polyps. In addition to colorectal lesions, several of these afflictions are associated with manifestations in other tissues including soft tissue tumors, osteomas, dental abnormalities, congenital hypertrophy of the retinal pigment epithelium (CHRPE), and cancers of the thyroid, small intestine, stomach, and brain (see, for example, Giardiello, F.M. (1995) "Gastrointestinal Polyposis syndromes and hereditary nonpolyposis colorectal cancer" in Rustgi AK (eds.) *Gastrointestinal Cancers: Biology, Diagnosis, and Therapy.* Philadelphia, Lippincott-Raven, pg. 367-377; Hamilton, S.R. et al. (1995) "The molecular basis of Turcot's syndrome." *New England J. Medicine* 332:839)). In contrast, patients with hereditary nonpolyposis colorectal cancer (HNPCC) lack an increase in the number of precursor adenomas yet share an increased risk for other ailments including cancer of the uterus, ovary and brain.

Breakthroughs in molecular biology have identified several of the genes/gene families involved in the initiation and progression of colon cancer (see, for example, Vogelstein, B. et al. (1988) "Genetic alterations during colorectal-tumor development." *New England J. Medicine* 319:525; Kinzler, K. and Vogelstein, B. (1998) "Colorectal Tumors" in The Genetic Basis of Human Cancer. McGraw-Hill). In addition to identifying genes involved in DNA mismatch-repair (hMSH1, hMSH2, HPMS1, hPMS2), cell growth (e.g. the oncogenes K-ras, H-ras or N-ras), and cell cycle regulation (e.g. tumor suppressors, p53), a growing body of evidence has suggested that a connection exists between cancer, cell adhesion and the Wnt/Wng pathway. Specifically attention has been drawn to a collection of gene products that include, but are not limited to, APC (adenomatous polyposis coli), β-catenin, TCF/LEF, GCK3, and cadherin.

Despite detailed knowledge of these and other genes involved in colorectal cancer, the art to date has not provided an efficient method for exploring the biological intricacies of colon cancer and identifying new putative therapeutic drugs for the prevention and treatment of this disease. Prophylactic colectomies are still routinely performed on FAP patients as a preferred method to reduce the risk of cancer and patients with metastatic disease usually receive radiation and/or current, broad-acting chemotherapeutic agents. Although such treatments can induce temporary remissions, they are often not curative, as evidenced by the fact that approximately 40% of the colon cancer patients die from the disease within 5 years. The present invention provides an opportunity to identify new drugs and drug targets that can be utilized to battle the increasing incidence of colon cancer that is predicted for the upcoming decade.

BRIEF SUMMARY OF THE INVENTION

The present invention provides methods for performing negative selections. In some embodiments, the negative selections are performed by introducing a genetic library into a population of target cells, collecting a subpopulation of cells that detach from a culturing surface (referred to herein as "floaters") and then recovering the introduced genetic material from that subpopulation. Other embodiments involve introducing a genetic library into a population of target cells, identifying cells that develop permeable membranes, and then recovering the transformed or transduced genetic material from that subpopulation. This later assay can be performed manually or in a roboticized, high-throughput format. In variations of these embodiments, cell-specific cytotoxic agents are identified by employing a counter-screening step wherein the genetic material from the subpopulation displaying a detachment, a membrane permeable, and/or other measures of a lethal phenotype is introduced into a second, different population of cells, and a second sublibrary of genetic material is obtained from a second subpopulation that does not display detachment and/or the lethal phenotype. For instance, cDNA sequences that induce cell death in i.e. a cancer cell line might be counter-screened against one or more normal cell types to identify cDNAs that kill the diseased cell type, but not the normal cell type. In still other embodiments, the invention provides methods for recovering agents that induce a lethal phenotype from dead and/or dying cells.

A variety of particular embodiments exist for each of these basic embodiments. In some particular embodiments, the lethal phenotype of the methodology may be apoptosis, necrosis, or growth arrest. In embodiments in which the lethal phenotype is apoptosis, the property of disattachment from a culturing substrate or membrane permeability may be used as a surrogate for apoptosis, thereby providing a technique for enriching the apoptotic cell population. In other particular embodiments, the genetic material may be partially sequenced, or the method steps may be reiterated in a second population of the same cells to further enrich for desirable cytotoxic/cytostatic sequences.

In some embodiments the cells may be mammalian cells, or more particularly primary cells, especially primary cells derived from epithelial or endothelial cells, stem cells, mesenchymal cells, fibroblasts, neuronal cells, or hematopoeitic cells. The mammalian cells may also be cancer cells, or more particularly cancer cells that are metastatic or derived from solid tumors. The cancer cells may particularly be derived from breast, colon, lung,

melanoma, brain, or prostate tissue. In other particular embodiments, the mammalian cells are genetically altered, and more particularly may be immortalized or transformed.

In embodiments that utilize the property of detachment of cells from a culturing surface, particular embodiments will feature a low background of spontaneously detaching cells, which may more particularly be no more than about 10%, or alternatively no more than about 2%. Cells having such low backgrounds include SW620 and HT29 colon cancer cells, T47D breast cancer cells, HuVEC cells, and others. In particular embodiments, the detaching or disadhering cells are collected over a period of at least about 12 hours.

The invention also lends itself to embodiments that screen for conditional cytotoxicity, wherein a genetic library is introduced into a population of target cells, exposing those target cells to a subtoxic threshold dose of a secondary reagent, collecting a subpopulation of cells displaying a lethal phenotype, and recovering genetic material from that subpopulation. Again, in particular embodiments the lethal phenotype may be apoptosis, necrosis or growth arrest. In other particular embodiments, the secondary reagent may be UV, X-ray or neutron radiation, or may be a chemotherapeutic agent, more particularly methotrexate, cisplatin, 5-fluorouracil, colchicines, vinblastine, vincristine, doxyrubicin or taxol. Particular embodiments include cancer cells, more particularly solid tumors, as target cells, counterscreening with a second cytotoxic substance, preconditioning the target cells prior to exposure with, e.g., growth factors, cytokines, chemokines, or activation of oncogenes. In an additional embodiment the assays described above can be performed in a high-throughput procedure whereby robotic elements are utilized to screen through thousands, or tens of thousands of potential cytotoxic agents for those sequences that induce cell death and/or cytostasis in the host cell of choice.

The invention also describes the criteria by which all compositions of matter isolated from such a screen are defined. These parameters include, but are not limited to, chemical makeup, size, fragments of such agents, binding properties, coding sequences, and more. The invention also lends itself to embodiments that define the construction or use of genetic libraries. Such libraries can be derived from natural sources such as genomic DNA or cDNA taken from human, mouse, nematode, fly, yeast, or other organisms, or they may be synthetically constructed using art-proven technologies.

The invention also encompasses the identification of small organic molecules that induce a lethal phenotype. In some embodiments, organic molecules that displace a proteinaceous cytotoxic agent from an endogenous protein target are obtained. In other

embodiments, organic molecules having a structure-activity relationship with that proteinaceous cytotoxic agent are identified.

The invention also provide a method of inducing cytotoxicity in a cell comprising the step of contacting said cell with an amount of a peptide fragment effective to induce cytotoxicity, said peptide fragment selected from the group consisting of: SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, and SEQ ID NO: 13. The invention further provides a method of inducing cytotoxicity in a cell comprising the step of introducing into said cell a polynucleotide encoding a peptide fragment in an amount effective to induce cytotoxicity, said peptide fragment selected from the group consisting of: SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, and SEQ ID NO: 13.

The invention also includes peptide fragments consisting of the amino acid sequences selected from the group consisting of SEQ ID NOS: 4 through 13 and the polynucleotides which encode said peptide fragments.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure Legends

Figure 1. Floater Screen. HT29 adenocarcinoma cells transfected with a perturbagen library are cultured for 3-5 days. Subsequently, floating cells are centrifuged, processed for genomic DNA, and PCR amplified to retrieve perturbagen encoding sequences. As one alternative to these procedures, floating cells can be stained with propidium iodide (PI) and sorted for PI+ (dead) cells. The PCR product is then ligated into a retroviral vector, packaged, and reintroduced into a naïve population of HT29 cells for additional rounds of screening and enrichment.

Figure 2. Perturbagen Disruption of Macromolecular Structures. Assembly of macromolecular subunits into stable quaternary structures requires the interaction between critical epitopes of the participating macromolecules. For instance, to maintain a helical structure composed of two heterologous subunits (α and β), α-α, β-β and α-β interactions must occur. Although many peptides can be found which show affinity to α or β subunits, most do not disrupt macromolecular assembly. One method of action, a perturbagen can bind to a critical epitope and disrupt the association of the two interacting proteins. In this

example, a small peptide (represented by the black triangle) binds to the α subunit and disrupts the interaction α and β subunits. As a result, the helix is disrupted.

Figure 3. A. Mapping the Biologically important region of a perturbagen. Four perturbagens are derived from different breakpoints within the same gene. By mapping the smallest sequence that is common to all four perturbagens (dotted line) it is possible to identify biologically critical regions (black box). B. Critical regions of a gene can be determined by deletion analysis. For instance, a series of N-terminal deletions (dotted line) can be tested for biological activity. In this example, full activity requires a molecule that is longer than deletion 2 but smaller than deletion 1.

Figure 4. Basic two-hybrid methodology. When bait and prey molecules interact, the Gal4-AD and Gal40-BD binding domains of the Gal4 transcriptional activator are reconstituted. As a result, this function unit can associate with the Gal1 UAS and induce transcription of the reporter gene (Leu2).

Figure 5. LANCE™. In the homogeneous assay, a Cy5 labeled perturbagen binds to an Eu-Target molecule in solution. A. When the two molecules are in close proximity, the emissions of the lanthanide chelate can excite Cy5 and give rise to a robust signal. B. In the presence of a small molecule inhibitor, the Cy5-perturbagen-Target-Eu interaction is prevented. Subsequent excitation of Eu results in little or no signal.

Figure 6. DELFIA™. In the heterogeneous assay, the target is immobilized to a solid support using an Eu labeled monoclonal antibody. Following incubation with the Cy5 labeled perturbagen, the well is washed to remove unbound Cy5. Due to the close proximity of the Eu and Cy5 moieties in the bound complex, excitation of the lanthanide chelate leads to excitation (and emission) of Cy5. In the presence of a small molecule inhibitor (black circles), the Eu-target and Cy5-perturbagen moieties never come in close proximity. In subsequent washes, the free, unbound, Cy5-peptide conjugate is removed and the Eu induced Cy5 signal is insignificant.

Figure 7. Diagram of the retroviral vector, pVT340.

Figure 8. Floater Enrichment. Bar graph showing the increased numbers of floaters over the course of seven cycles of enrichment.

Figure 9. Sequence of BID (SEQ ID NO: 3). Arrows indicate positions of the two N-terminal breakpoints.

Figure 10. Peptide sequences (SEQ ID NOS: 4-11) of additional perturbagens.

Figure 11. Alignment of Amino Acid Sequence of 1F1 perturbagen (SEQ ID NO: 13) and BC002905 (SEQ ID NO: 12). Bold, underlined letters represent amino acid changes in 1F1 clone.

Figure 12. Kill Indexes of Perturbagens in HT29 and SW620 Cells.

5 **Figure 13.** Kill Indexes of Perturbagens In-Frame (IF) and Out-Of-Frame (OF).

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Agents isolated from the methods described herein have broad potential and application. For example, each proteinaceous agent (or a mimetic thereof identified through,
10 e.g., routine small molecule screens) may be useful as a direct therapeutic agent in the treatment of a variety of forms of cancer. With each such new agent, a corresponding target molecule can be readily identified using standard interaction methodologies such as the two-hybrid technique. Such targets are useful in the development of novel drugs for new chemotherapeutic strategies and may provide useful diagnostic tools for profiling the genetic
15 background (genotype) of the particular disease under study.

A. Overview of the Invention

The invention describes the isolation of new and previously unidentified agent(s) that alter a cell's viability. These agents (referred to herein as "perturbagens") consist of polypeptides and may be encoded by a naturally derived library of compounds such
20 as a cDNA or genomic DNA (gDNA) expression library, or an artificial library comprising synthetic oligonucleotide sequences of a desired length or range of lengths, e.g. a random peptide library.

The perturbagens described herein were isolated using a phenotypic assay referred to as a "floater assay" (See priority documents U.S. Patent No. 5,955,275, "Methods
25 for identifying nucleic acid sequences encoding agents that affect cellular phenotypes," and USSN 09/504,132, the disclosures of which are incorporated by reference herein in its entirety). Briefly, the floater assay identifies agents that induce a lethal phenotype. To accomplish this, a population of polynucleotide sequences (a "library") is generated using a variety of techniques familiar to the art. After ligating this material into a standard
30 expression vector, the library is transferred into a population of cells that exhibit strong adherence to a solid support (e.g. plastic, agarose) and screened for sequences that induce cells to lose their adhesion properties (see Figure 1). Previous studies have demonstrated that

a strong correlation exists between loss of adhesion (i.e. floating) and cell death. Thus, by collecting cells that are released from the surface, the assay advantageously identifies one or more relevant sequences from the library that induces the desired phenotype, cell death. Dead and/or dying cells are separated from the rest of the population by either i) collecting the media that overlays the adherent cell monolayer, or ii) staining the entire population of cells (floating + adherents) with one of several dyes/markers that distinguishes dead from vital cells (e.g. propidium iodide, Apo2.7 antibody, annexin) and separating the desired population by Fluorescent Activated Cell Sorting (FACS). Alternatively, dead cells can be sorted using the Forward Scatter/Side Scatter option of the FACS machine which separates cells on the basis of size and granularity. Previous studies have shown that in cultures containing both live and dead cells, two populations (referred to as Pop1 and Pop2) are easily distinguished. Cells which are large and (in general) lightly granulated, have been found to be healthy (i.e. alive) and fall into the Pop1 population. In contrast, apoptotic cells are typically smaller and highly granulated, and fall into the Pop2 population. Regardless of whether the dead cells are stained or distinguished by their size/refractive properties, FACS machines are both highly sensitive and efficient (obtaining screening speeds of approximately 10,000 to approximately 65,000 cells or more per minute) thus facilitating identification of biologically relevant sequences that exist at low frequencies within a cell population. Subsequence PCR amplification, sublibrary formation, and re-screening of the perturbagen encoding sequences derived from the dead cell population enables further enrichment of sequences that induce cell death.

Perturbagen identification may elucidate the function of known genes, or alternatively may work in a "black-box" approach to identify new genes, gene products, or cellular targets. Thus, in some instances, perturbagens may be encoded by a previously identified gene (or gene fragment thereof). Such a gene may be one whose contribution to the disease pathway has previously been identified. Alternatively, a particular gene product may have been previously identified, yet its contribution to a given pathway or its phenotype may have been previously unrecognized. In cases where the perturbagen is a fragment of a protein, the observed phenotype may be not be endogenous to the intact gene product, but instead, is the result of said fragment (e.g. the C-terminal) being separated from, for instance, a regulatory region that is normally present in the intact protein but has been omitted during the library formation. In yet other cases, the perturbagen may be found to have no homology with any previously identified polynucleotide or proteinaceous agent. Such perturbagens may be derived from previously unidentified genes, or alternatively, may be derived from

random sequences that have the proper conformation and/or chemical characteristics needed to alter or modulate one or more components of a pathway(s) that influences the phenotype under investigation. In the methodology described herein, no prior knowledge of the perturbagen or of its corresponding gene, gene product or cellular target is necessary.
5 Moreover, because it is possible for multiple perturbagens to assume similar secondary or tertiary conformations and/or have shared or related chemistries, two or more variants of the same perturbagen may be identified and isolated from a single library without any additional screening steps. Thus, one need not spend laborious hours designing, redesigning, or manipulating any "candidate molecules", and thus does not bias the experiment with
10 preconceived conceptions of what will or will not induce the phenotype of interest.

B. Phenotypic Probes

The invention encompasses both the phenotypic probes (perturbagens) described herewith and the polynucleotide sequences encoding them. In this context, the term "perturbagen" or "phenotypic probe" refers to any compound that is proteinaceous in
15 nature and is, through its interaction with specific cellular target(s) or other such component(s), capable of disrupting or activating a particular signaling pathway and/or cellular event. As one of ordinary skill appreciates, agents may be described by their RNA sequence, amino acid sequence, or correlative DNA sequence. Alternatively, the agents can be sufficiently described in terms of their identity as isolates of a library that exhibit a
20 particular biological activity.

Perturbagens may be encoded by a variety of genetic libraries, including those developed from cDNA, gDNA, and random, synthetic oligonucleotides synthesized using current available methods in chemistry (see, for example, Caponigro et al. (1998) "Transdominant genetic analysis of a growth control pathway." *PNAS* 95:7508-7513;
25 Caruthers, M.H. et al. (1980) Nucleic Acids Symposium, Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symposium, Ser. 7:225-232; Cwirla, S.E. et al. (1990) "Peptides on phage: a vast library of peptides for identifying ligands." *Proc Natl Acad Sci* 87(16):6378-82). Alternatively, the perturbagen itself, or fragments of the perturbagen, can be synthesized using chemical methods. For example, peptide and RNA synthesis can be performed using
30 various techniques (Roberge, J.Y. et al. (1995) "A strategy for a convergent synthesis of N-linked glycopeptides on a solid support." *Science* 269:202-204; Zhang, X. et al. (1997) "RNA synthesis using a universal base-stable allyl linker." *NAR* 25(20):3980-3983).

Automated synthesis may be achieved using commercially available equipment such as the ABI 431A peptide synthesizer (Perkin-Elmer).

In some cases, such as this one, the polynucleotide sequence encoding a perturbagen represents a fragment of an existing gene. Using currently available software, it is possible to identify the full length cDNA by aligning the perturbagen encoding sequence with pre-existing sequences maintained in, for instance, publicly available genomic and/or EST data bases. In situations where the gene has not been identified, the perturbagen can be readily used to "reverse engineer" and identify the gene from which the phenotypic probe is derived. In this context, the term "gene" includes both the coding and antisense strands, the 5' and 3' regions that are not transcribed but serve as transcriptional control domains, and transcribed but not expressed domains such as introns (including splice junctions), polyadenylation signals, translation initiation signals, and the like.

In the case where a perturbagen is encoded by only a portion of a particular gene, the nucleic acid sequence of such a perturbagen may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences. One such method, restriction site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector (Sarkar, G. (1993) "Restriction-site PCR: a direct method of unknown sequence retrieval adjacent to a known locus by using universal primers." *PCR Methods Applic.* 2:318-322). Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences (see Trigilia, T. et al. (1988) "A procedure for in vitro amplification of DNA segments that lie outside the boundaries of known sequences." *NAR.* 16:8186). A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) "Capture PCR: efficient amplification of DNA fragments adjacent to a known sequence in human and YAC DNA." *PCR Methods Applic.* 1:111-119). In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double stranded sequence into a region of known sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art (Parker, J.D. et al. (1991) "Targeted gene walking polymerase chain reaction." *NAR.* 19:3055-3060). In addition, one may use nested primers and PROMOTERFINDER libraries (Clontech, Palo Alto, CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions.

For all PCR based methods, primers may be designed, using commercially available software such as OLIGO 4.06 Primer Analysis software (National Biosciences, Plymouth, MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

5 In one particular embodiment, the invention encompasses proteinaceous perturbagens, biologically active fragments, (N-terminal, C-terminal, or internal) or variants thereof. The term "proteinaceous perturbagens" encompasses peptides, oligo- or polypeptides, proteins, protein fragments, or protein variants. Some proteinaceous perturbagens can be as short as three amino acids in length. Alternatively, these agents can 10 be greater than 3 amino acids but less than ten amino acids. Other agents can be greater than ten amino acids but shorter than 30 amino acids in length. Still other agents can be greater than 30 amino acids but less than 100 amino acids in length. Still other agents can be greater than 100 amino acids in length. Naturally occurring proteinaceous perturbagens (i.e. those derived from cDNA or genomic DNA) exhibit a range in size from as little as three to several 15 hundred amino acids. In contrast, synthetic perturbagens (such as those present in a synthetic peptide library) may range in size from three amino acids to fifty amino acids in length and more preferably, from three to 20 amino acids in length, and yet more preferably, about 15 amino acids in length.

Proteinaceous perturbagens can exert their effects by multiple means. For 20 example, a peptide may act by binding and disrupting the interactions between two or more proteinaceous entities within the cell (see Figure 2). In another instance, perturbagens action can result from an agent having a particular enzymatic activity and expressing that activity in, for instance, i) an unregulated fashion, or ii) in a novel compartment. Alternatively, a peptide perturbagen can bind to, and disrupt translation of a particular mRNA molecule. As still 25 another alternative, peptide perturbagens may bind to genomic DNA and disrupt gene expression by altering the ability of one or more transcription factor(s) (e.g. activators or repressors) from binding to a critical enhancer/promoter region of the regulatory region of the gene.

Penetrance is another property of perturbagens. Penetrance is defined as the 30 number of cells exhibiting a particular phenotype divided by the total number of cells in the experiment (when a perturbagens is present in the cells), minus the total number of cells exhibiting a particular phenotype divided by the total number of cells in the experiment when the perturbagens is not present in the cells. The penetrance of any given perturbagen can vary depending upon a variety of parameters including 1) the cell type it is being expressed

in, 2) the vector being used to express the perturbagen, 3) the biological stability (half-life) of the perturbagen or mRNA encoding the perturbagen 4) the concentration of the perturbagen in the cell, as well as other parameters. In addition, the penetrance of a given perturbagen may not be directly related to the "quality" or "usefulness" of the perturbagen molecule. Thus, although penetrance is a factor that impacts how immediately a given perturbagen can be seen to exert an effect, in some instances, a desirable, biologically active perturbagen may present a relatively low rate of penetrance. Furthermore, the penetrance of a given perturbagen may not be directly related to the "quality" of the molecular target it identifies.

As one of ordinary skill will appreciate, perturbagens of low penetrance may be obtained and manipulated via standard cycling and/or amplification procedures. Thus, some preferred perturbagens may exhibit as low as 1-2% penetrance. Other preferred perturbagens may exhibit between 2% and 5% penetrance, between 5 and 10% penetrance, 10% and 20% penetrance, between 20% and 50% penetrance, or even in some instances, between 50% and 100% penetrance.

C. Sequence Variants

In another embodiment, the invention includes sequence variants of both the phenotypic probes and the polynucleotide sequences that encode them. In this context the term "variant" refers to biologically active forms of the perturbagen sequence (or the polynucleotide sequence that encodes the perturbagen) that differ from the sequence of the initial perturbagen. Thus, in the case of proteinaceous perturbagens, variants contain at least one amino acid substitution, deletion, or insertion from the original isolated form of the perturbagen that provides biological properties that are substantially similar to those of the initial perturbagen. Similarly, variants may occur in the RNA/DNA that encodes each phenotypic probe and thus it is possible for each to contain at least one nucleotide substitution, deletion, or insertion when compared to the original isolated sequence.

In addition to being described by their respective sequence, variants may also be identified by the relative amounts of homology they have in common with the original perturbagen sequence. "Homology" is defined as the percentage of residues in a candidate sequence that are identical with the residues in the reference sequence after aligning the two sequences and introducing gaps, if necessary, to achieve the maximum percent of overlap (see, for example, Altschul, S.F. et al. (1990) "Basic local alignment search tool." *J Mol Biol* 215(3):403-10; Altschul, S.F. et al. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." *Nucleic Acids Res* 25(17):3389-402).

Alternatively, a variant of a proteinaceous perturbagen may be described in terms of the nature of an amino acid substitution. "Conservative" substitutions are those in which the substituting residue is structurally or functionally similar to the substituted residue. In non-conservative substitutions, the substituting and substituted residue will be from structurally or functionally different classes. For the purposes herein, these classes are as follows: 1. Electrospective: R, K, H; 2. Electronegative: D, E; 3. Aliphatic: V, L, I, M; 4. Aromatic: F, Y, W; 5. Small: A, S, T, G, P, C; 6. Charged: R, K, D, E, H; 7. Polar: S, T, Q, N, Y, H, W; and Small Hydrophilic: C, S, T. Interclass substitutions generally are characterized as nonconservative, while intraclass substitutions are considered to be conservative. In some instances, variant polypeptides sequences can have 65-75% homology with the original agent. In other embodiments, variants have between 75% and 85% homology with the original agent. In still other embodiments, variants will have between 85% and 95% homology with the original perturbagen agent. In yet other embodiments, variants have between 95% and greater than 99% polypeptide sequence identity with the original perturbagen agent. In some cases, the homology between two perturbagens (variants) is confined to a small region of the molecule (e.g. a motif). Such conserved sequences are often indicative of regions that contain biologically important functions and suggest the perturbagens share a common cellular target. In these situations, while only limited and conservative amino acid changes are desirable within the region of the motif, greater levels of variation can exist in adjacent and more distal portions of the polypeptide.

The RNA encoding each perturbagen may also be described in terms of percent homology. In some instances, the variant ribonucleotide sequences can have 65-75% homology with the original agent. In other embodiments, the variants have between 75% and 85% homology with the original agent or between 85% and 95% homology with the original perturbagen sequence, or even between 95% and greater than 99% sequence identity with the original perturbagen agent. Again, greater variation can, in some embodiments, exist outside an identified region/motif without altering biological activity.

Lastly, in reference to the DNA sequences encoding proteinaceous perturbagens, one who is skilled in the art will appreciate that the degree of variance will depend upon and/or reflect the degeneracy of the genetic code. As one in the art appreciates, a given protein sequence is equivalently encoded by a large number of polynucleotide sequences. Therefore, the invention encompasses each variation of polynucleotide sequence that encodes the given perturbagen, such variations being made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of each perturbagen.

For each proteinaceous perturbagen described by amino acid sequence herein, all such corresponding DNA variations are to be considered as being specifically disclosed.

Variants of phenotypic probes may arise by a variety of means. Some variants may be artifactual and result from, for instance, errors that occur in the process of PCR amplification or cloning of the perturbagen encoding sequence. Alternatively, variants may be constructed intentionally. For instance, it may be advantageous to produce nucleotide sequences encoding perturbagens possessing a substantially different codon usage. Codons may be selected to increase the rate at which expression of the peptide or RNA occurs in a particular prokaryotic or eukaryotic cell in accordance with the frequency with which particular codons are utilized by the host (Berg, O.G. (1997) "Growth rate-optimized tRNA abundance and codon usage." *J Mol Biol* 270(4):544-50). Additional reasons for substantially altering the nucleotide sequence encoding proteinaceous perturbagens (without altering the encoded amino acid sequences) include, but are not limited to, producing RNA transcripts that have increased half-life. This may be accomplished by altering a sequence's structural stability (see, for example, Gross, G. et al. (1990) "RNA primary sequence or secondary structure in the translational initiation reign controls expression of two variant interferon-beta genes in Escherichia coli." *J Biol Chem.* 265(29):17627-36; Ralston, C.Y. et al. (2000) "Stability and cooperativity of individual tertiary contacts in RNA revealed through chemical denaturation." *Nat Struct Biol.* 7(5):371-4), or through addition of untranslated sequences that increase RNA stability/half-life through RNA-protein interactions (see, for example, Wang, W. et al. (2000) "HuR regulates cyclin A and cyclin B1 mRNA stability during cell proliferation." *EMBO J.* 19(10):2340-50; Shetty, S. and Idell, S. (2000) "Posttranscriptional regulation of plasminogen activator inhibitor-1 in human lung carcinoma cells in vitro." *Am J Physiol Lunch Cell Mol Physiol* 278(1):L148-56). Also included the category of "intentional variants" are those whose sequence has been altered in order to add or deleted sites involved in post-translational modification. Included in this list are variants in which phosphorylation sites, acetylation sites, methylation sites, and/or glycosylation sites have been added or deleted (see, for example, Wicker-Planquart, C. (1999) "Site-directed removal of N-glycosylation sites in human gastric lipase." *Eur J Biochem.* 262(3):644-51; Dou, Y. (1999) "Phos-phorylation of linker histone H1 regulates gene expression in vivo by mimicking H1 removal." *Mol Cell.* 4(4):641-7).

Variants may also arise as a result of simple and relatively routine techniques involving random mutagenesis or "DNA shuffling"; procedures that are often used to rapidly evolve perturbagen encoding sequences and allow identification of variants that have

increased biological stability or activity (see, for instance, Ner, S.S. et al. (1988) "A simple and efficient procedure for generating random point mutations and for codon replacements using mixed oligonucleotides." *DNA* 7:127-134; Stemmer, W. (1994) "Rapid evolution of a protein in vitro by DNA shuffling." *Nature* 370:389-391). For instance, in mutagenic PCR, the fragment encoding the perturbagen is PCR amplified under conditions that increase the error rate of *Taq* polymerase. This may be accomplished by i) increasing the $MgCl_2$ concentrations to stabilize non-complementary pairings, ii) addition of $MnCl_2$ to diminish template specificity of the polymerase and iii) increasing the concentration of dCTP and dTTP to promote misincorporation of basepairs in the reaction. As a result of this process, the error rate of *Taq* polymerase may be increased from 1.0×10^{-4} errors per nucleotide per pass of the polymerase, to approximately 7×10^{-3} errors per nucleotide per pass. Amplifying a perturbagen-encoding sequence under these conditions allows the development of a library of dissimilar sequences which can subsequently be screened for variants that exhibit improved biological activity.

In addition to variants that are created by artificial or accidental means, natural variants may also exist. For instance, in the course of screening any given genomic or cDNA library, it is possible that a perturbagen, derived from a sequence that exists in multiple copies within the genome (e.g. duplications, repetitive sequences), may be isolated numerous times. Such sequences often contain polymorphisms that result in alterations in the encoded RNA and polypeptide sequence (see, for example, Satoh, H. et al. (1999) "Molecular cloning and characterization of two sets of alpha-theta genes in the rat alpha-like globin gene cluster." *Gene* 230(1):91-9) and thus, may represent natural variants of the perturbagen agent. Alternatively, if multiple libraries are utilized to screen for perturbagens and two or more of those libraries are derived from unrelated individuals or tissues, it is possible that variants may be isolated as a result of allelic or splice variation respectively (see, for example, Posnett, D.N. (1990) "Allelic variations of human TCR V gene products." *Immunol Today.* 11(10):368-73). Variants of phenotypic probes may arise by these and other means.

D. Biologically Active Fragments

Some embodiments of the invention encompass "biologically active fragments" of a given proteinaceous perturbagen. In this context, the term "fragment" refers to any portion of a proteinaceous perturbagen that is at least 3 amino acids in length, and the corresponding nucleic acids that encode such fragments. The terms "biologically relevant" or "biologically active" refer to that portion of a protein or protein fragment, RNA or RNA

fragment, or DNA fragment that encodes either of the two previous entities, that is responsible for an observable phenotype (or for activation of a correlative reporter construct). Thus, biologically active fragments may be comprised of N-terminal, C-terminal, or internal fragments of peptide perturbagens. In some instances, the fragment encodes or represents portions of a natural gene. In other instances the fragment is derived from a larger polynucleotide or polypeptide that has no known natural counterpart. In still other instances, biologically active regions of a perturbagen can be artificially synthesized (by chemical or recombinant methods) so that multiple, tandem copies of the phenotypic probe are covalently linked together and expressed. All such biologically active perturbagen fragments are, in turn, encoded by a variety of correlative DNA sequences.

The biologically active portion of a molecule can be identified by several means. In some instances, biological relevant regions can be deduced by simple physical mapping of families of overlapping sequences isolated from a phenotypic assay (Hingorani, K. et al. (2000) "Mapping the functional domains of nucleolar protein B23." *J Biol Chem* May 26). For instance, in the course of any given screen, multiple perturbagens, derived from alternative breakpoints of the same gene, may be isolated from one or more genetic libraries (see Figure 3). The smallest region that is common to all of the perturbagens can demarcate the area of biological importance.

Alternatively, critical regions of a perturbagen can frequently be distinguished by comparing the polynucleotide and/or amino acid sequence of two or more perturbagens that share a common target (see, for example, Grundy, W.N. (1998) "Homology detection via family pair-wise search." *J Comput Biol.* 5(3):479-9; Gorodkin, J. et al. (1997) "Finding common sequence and structure motifs in a set of RNA sequences." *Ismb* 5:120-3). In this instance, conserved sequences (or motifs) that are identified by this form of analysis often provide important clues necessary to determine biologically important regions of a given molecule. Alternatively, methods that identify biologically relevant regions by altering or deleting regions of the perturbagen molecule can also be used. For instance, the gene encoding a particular perturbagen can be subjected to deletion analysis whereby portions of the gene are removed in a systematic fashion, thus allowing the remaining entity to be retested for its ability to evoke a biological response (see Figure 3 and Huhn, J. et al. (2000) "Molecular analysis of CD26-mediated signal transduction in cells." *Immunol Lett* 72(2):127-132; Davezac, N. et al. (2000) "Regulation of CDC25B phosphatases subcellular localization." *Oncogene* 19(18):2179-85).

Alternatively, biologically critical regions of a molecule can be identified by inducing mutations in the sequence encoding the polypeptide (see, for example, Ito, Y. et al. (1999) "Analysis of functions regions of YPM, a superantigen derived from gram-negative bacteria." *Eur J Biochem*; 263(2):326-37; Kim, S.W. et al. (2000) "Identification of functionally important amino acid residues within the C2-domain of human factor V using alanine-scanning mutagenesis." *Biochemistry* 39(8):1951-8.). Subsequent testing of the variants of said molecule for biological activity enables the investigator to identify regions of the perturbagen that are both critical and sensitive to manipulation. Molecular probes such as monoclonal antibodies and epitope-specific peptides can also be useful in the identification of biologically important regions of a perturbagen (see, for example, Midgley, C.A. et al. (2000) "An N-terminal p14ARF peptide blocks Mdm2-dependent ubiquitination in vitro and can activate p53 in vivo." *Oncogene* 19(19):2312-23; Lu, D. et al. (2000) "Identification of the residues in the extracellular regions of KDR important for interaction with vascular endothelial growth factor and neutralizing anti-KDR antibodies." *J Biol Chem* 275(19):14321-30). In this procedure, probes that bind and thus mask specific regions of a perturbagen can be tested for their ability to block the biological activity of the molecule. These techniques (as well as others) can be used to map the boundaries of any given biologically active residues.

E. Heterologous Sequence

In another embodiment, the invention encompasses all heterologous forms of the phenotypic probes and the polynucleotide sequences encoding them described herewith. In this context, "heterologous sequence(s)" include versions of the perturbagens that are i) scaffolded by other entities, ii) tagged with marker sequences that can be recognized by antibodies or specific peptides, or iii) altered to transform post-translational patterns of modification.

1. Scaffolds

The terms "scaffold" refers to a proteinaceous or RNA sequence to which the perturbagen or perturbagen encoding ribonucleic acid sequence is covalently linked during synthesis to provide e.g., conformational stability and/or protection from degradation. Thus, peptide perturbagens can be fused to protein scaffolds at N-terminal, C-terminal, or internal sites. Similarly, the RNA sequences encoding those perturbagens can be fused to RNA sequences at 5', 3' or internal sites and increase the stability of the messenger RNA (mRNA) of said agent.

In some instances, scaffolds may be a relatively inert protein, (i.e. having no enzymatic activity or fluorescent properties). As one example of an inert scaffold, we created a non-fluorescent variant of GFP called dead-GFP (also referred to as "dGFP"). In this case, the nonfluorescent variant was brought about by conversion of Tyr → Phe at codon 66 of EGFP. Such proteins can be stably expressed in a wide variety of cell types without disrupting the normal physiological functions of the cell. In addition, such chimeric fusions can easily be detected by Western Blot analysis using antibodies directed against GFP and are useful in determination of intracellular expression levels of perturbagens. In other instances, scaffolds may serve a dual function, e.g., increasing perturbagen stability while the at the same time, serving as an indicator or gauge of the level of perturbagen expression. In this case, the scaffold may be an autofluorescent molecule such as a green fluorescent protein (GFP, Clontech) or embody an enzymatic activity capable of altering a substrate in such a way that it can be detected by eye or instrumentation (e.g. β galactosidase). Lastly, by modifying the perturbagen sequences or the scaffold to which they are attached with various "localization" signals, the perturbagen may be directed to a particular compartment within the host cell. For example, proteinaceous perturbagens can be directed to the nucleus of certain cell types by attachment of a nuclear localization sequence (NLS); a heterogeneous sequence made up of short stretches of basic amino acid residues recognized by importins alpha and/or beta (see, for example, Lobl, T.J. et al. (1990) SV40 large T-antigen nuclear signal analogues: successful nuclear targeting with bovine serum albumin but not low molecular weight fluorescent conjugates." *Biopolymers* 29(1):197-203).

2. Antibodies and Antibody-Tagged Perturbagens

Perturbagens can be constructed to contain a heterologous moiety (a "tag") that is recognized by a commercially available antibody. Such heterologous forms may facilitate studies of subjects including, but not limited to, i) perturbagen subcellular localization, ii) intracellular concentration assessment and iii) target binding interactions. In addition, the tagging of a perturbagen may also facilitate purification of fusion proteins using commercially available matrices (see, for example, James, E.A. et al. "Production and characterization of biologically active human GM-CSF secreted by genetically modified plant cells." *Protein Expr Purif.* 19(1):131-8; Kilic, F. and Rudnick, G. (2000) "Oligomerization of serotonin transporter and its functional consequences." *Proc Natl Acad Sci U S A.* 97(7):3106-11). Such moieties include, but are not limited to glutathione-S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, c-myc, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable

purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, c-myc and HA enable immuoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. Such fusion proteins may also be engineered to contain a proteolytic cleavage site located between the perturbagen sequence and the heterologous protein sequence, so that the perturbagen may be cleaved away from the heterologous moiety following purification. A variety of commercially available kits may be used to facilitate expression and purification of fusion proteins.

An additional embodiment of the invention includes antibodies that recognize the perturbagen itself or cellular targets of the perturbagen. Antibodies directed against perturbagens or cellular targets may be useful for a variety of purposes including i) therapeutics, ii) diagnostic assays iii) immunocytochemistry, iv) target identification, and v) purification. Such reagents may include, but are not limited to, polyclonal, monoclonal, chimeric, and single antibodies, Fab fragments, and fragments produced by a Fab expression library. For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans and others may be immunized by injection with a perturbagen or any fragment thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to Freund's, mineral gels such as aluminum hydroxide, and surface-active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

Monoclonal antibodies that recognize perturbagens may be prepared using any technique that provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV hybridoma technique. (see, for example, Kohler, G. et al. (1975) "Continuous cultures of fused cells secreting antibody of predefined specificity." *Nature* 256:495-497; Kozbor, D. et al. (1985) "Specific immunoglobulin production and enhanced tumorigenicity following ascites growth of human hybridomas." J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) *PNAS* 80:2026-2030; and Cole, S.P. et al. (1984) "Generation of human monoclonal antibodies reactive with cellular antigens" *Mol. Cell Biol.* 62:109-120).

In addition, techniques developed for the production of chimeric antibodies, such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. See, e.g., Morrison, S.L. et al. (1984) "Chimeric human antibody molecules: mouse antigen-binding domains with human constant region domains." *PNAS* 81:6851-6855); Neuberger, M.S. et al. (1984) "Recombinant antibodies possessing novel effector functions." *Nature* 312:604-608; and Takeda, S. et al. (1985) "Construction of chimeric processed immunoglobulin genes containing mouse variable and human constant region sequences." *Nature* 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce perturbagen-specific antibodies (see, e.g. Burton, D.R. (1991) "A large array of human monoclonal antibodies to type 1 human immunodeficiency virus from combinatorial libraries of asymptomatic seropositive individuals." *PNAS* 88:10134-10137).

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (see, for example, Orlandi, R. et al. (1989) "Cloning immunoglobulin variable domains for expression by the polymerase chain reaction." *PNAS* 86:3833-3837; Winter, G. et al. (1991) "Man-made antibodies." *Nature* 349:293-299).

Antibody fragments that contain specific binding sites for perturbagens may also be generated. For example, such fragments include, but are not limited to, $F(ab)_2$ fragments produced by pepsin digesting of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the $F(ab)_2$ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, for example, Huse, W.D. et al. (1989) "Generation of a large combinatorial library of the immunoglobulin repertoire in phage lambda." *Science* 246:1275-1281).

3. Chemically Modified Perturbagens

In addition to the chimeric variants described above, chemical modification encompass a variety of modifications including, but not limited to, perturbagens that have been radiolabeled with ^{32}P or ^{35}S, acetylated, glycosylated, or labeled with fluorescent molecules such as FITC or rhodamine. These modifications may be directly imposed on the perturbagen itself (see, for example, Shuvaev, V.V. et al. (1999) "Glycation of apolipoprotein E impairs its binding to heparin: identification of the major glycation site." *Biochim Biophys*

Acta 1454(3):296-308; Dobransky, T. et al. (2000) "Expression, purification and characterization of recombinant human choline acetyltransferase: phosphorylation of the enzyme regulates catalytic activity." *Biochem J.* 349(Pt. 1):141-151). Alternatively, changes may be made to the polynucleotide sequence encoding the perturbagen so as to alter the pattern of phosphorylation, acetylation, or glycosylation.

F. Hybridization

The invention also encompasses polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences encoding phenotypic probes and said variants of such entities described previously, under various conditions of stringency. Such reagents may be useful in i) therapeutics, ii) diagnostic assays, iii) immunocytology, iv) target identification, and v) purification. For example, if the sequence encoding a particular perturbagen is introduced into a subject for gene therapeutic purposes, it may be necessary to monitor the success of integration and the levels of expression of said agent by Southern and Northern Blot analysis respectively (Pu, P. et al. (2000) "Inhibitory effect of antisense epidermal growth factor receptor RNA on the proliferation of rat C6 glioma cells in vitro and in vivo." *J Neurosurg.* 92(1):132-9). In other instances, hybridization may be used as a tool to define or describe a perturbagen variant or fragment, and a hybridizing sequence thus may have direct relevance as mimetic or other such therapeutic agent.

The term "hybridization" refers to any process by which a strand of nucleic acid binds with a complementary or near-complementary strand through base pairing. There are several parameters that play a role in determining whether two polynucleotide molecules will hybridize including salt concentrations, temperature, and the presence or absence of organic solvents. For instance stringent salt concentrations will ordinarily be less than about 750mM NaCl and 75mM trisodium citrate, preferably less than about 500mM NaCl and 50mM trisodium citrate, and most preferably less than about 250mM NaCl and 25mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent (e.g. formamide) while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and most preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30°C, more preferably of at least about 37°C, and most preferably of at least about 42°C. Varying additional parameters, such as hybridization time, the concentration of detergent and the inclusion or exclusion of carrier DNA are well known to those skilled in the art. Various levels of stringency are accomplished by combining these various conditions as needed. In a

preferred embodiment, hybridization will occur at 30°C in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37°C in 500 mM NaCl, 50mM trisodium citrate, 1% SDS, 35% formamide and 100ug/ml denatured salmon sperm DNA (ssDNA). In a most preferred embodiment, hybridization will occur at 5 42°C in 250 mM NaCl, 25mM trisodium citrate, 1% SDS, 50% formamide and 200ug/ml denatured ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

The washing steps that follow hybridization can also vary greatly in stringency. Wash stringency conditions can be defined by salt concentration and by 10 temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentrations for the wash steps will preferably be less than about 30mM NaCl and 3mM trisodium citrate, and most preferably less than about 15mM NaCl and 1.5mM trisodium citrate. Stringent temperature conditions for the wash steps will ordinarily include temperatures of at least about 25°C, more preferably 15 of at least about 42°C, and most preferably of at least about 68°C. In a preferred embodiment, wash steps will occur at 25°C in 30mM NaCl, 3mM trisodium citrate and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 42°C in 15mM NaCl, 1.5mM trisodium citrate and 0.1% SDS. In a most preferred embodiment, wash steps will occur at 68°C in 15mM NaCl, 1.5mM trisodium citrate and 0.1% SDS. Additional variations on these 20 conditions will be readily apparent to those skilled in the art.

G. Expression Vectors

The DNA sequence encoding each perturbagen (or variant or fragment thereof) may be inserted into an expression vector which contains the necessary elements for transcriptional/translational control in a selected host cell. The vector thus may include 25 regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions, mRNA stabilizing sequences or scaffolds, for optimal expression of the perturbagen in a given host. For instance, intracellular perturbagen levels can be modulated using alternative promoter sequences such as CMV, RSV, and SV40 promoters, to drive transcription (see, for example, Zarrin, A.A. et al. (1999) "Comparison of CMV, RSV, SV40 30 viral and V lambda 1 cellular promoters in B and T lymphoid and non-lymphoid cell lines. *Biochim Biophys Acta.* 1446(1-2):135-9). Alternatively, inducible promoter systems, (e.g. ponesterone-induced promoter (PIND, Invitrogen, see Dunlop, J. et al. (1999) "Steroid hormone-inducible expression of the GLT-1 subtype of high-affinity 1-glutamate transporter

in human embryonic kidney cells." *Biochem Biophys Res Commun.* 265(1):101-5), tissue specific enhancers (see Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162), or scaffolding molecules (see, for example, Abedi, M. et al. (1998), "Green fluorescent protein as a scaffold for intracellular presentation of peptides." *Nucleic Acid Research* 26(2):623-630) can be used to modulate intracellular perturbagen levels.

Specific initiation signals may be used to achieve more efficient translation of sequences encoding the perturbagen. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding the perturbagen and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence is inserted, exogenous translational control signals including an in-frame ATG initiation codon are provided by the vector. Such exogenous translational elements and initiation codons may be of various origins, both natural and synthetic.

In some instances, sequences that stabilize the RNA transcript or direct the RNA sequence to a particular compartment may be included (see, for instance, Wood Chuck post transcriptional regulatory element, WPRE, Zufferey, R. et al. (1999) "Woodchuck hepatitis virus posttranscriptional regulatory element enhances expression of transgenes delivered by retroviral vectors." *J Virol* 73(4):2886-92).

Methods which are well known to those skilled in the art are used to construct expression vectors containing sequences encoding the perturbagens and the appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination (see Ausubel, F.M. (1993) "Current Protocols in Molecular Biology", Wiley, John & Sons, Incorporated).

A variety of paired expression vector/host systems may be utilized to contain and express sequences encoding the perturbagens. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid or cosmid DNA expression vectors; yeast transformed with yeast expression vectors, insect cell systems infected with viral expression vectors (e.g. baculoviris), plant cell systems transformed with viral expression vectors (e.g. tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g. Ti or pBR322 plasmids; or animal systems). The host cell employed does not limit the invention.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding the perturbagens. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding perturbagens can be achieved using a multifunctional E. coli vector such as pBLUESCRIPT (Stratagene, La Jolla, CA). Ligation of sequences encoding perturbagens into the vector's cloning site disrupts the lacZ gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the clones sequence. (see e.g., Van Heeke, G. and Schuster, S.M. (1989) "Expression of human asparagine synthetase in Escherichia coli." *J. Biol. Chem.* 264:5503-5509). When large quantities of perturbagens are needed, e.g. for the production of antibodies, vectors which direct high level expression of perturbagens may be used. Exemplary vectors feature the strong, inducible T5 of T7 bacteriophage promoter.

Yeast expression systems may also be used for production of perturbagens. A number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or related strains. In addition, such vectors can be designed to direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences in the host genome for stable propagation. (see, e.g. Bitter, G.A. et al. (1987) "Expression and secretion vectors for yeast." *Methods Enzymology.* 153:516-544; and Scorer, C.A. et al. (1994) "Rapid selection using G418 of high copy number transformants of Pichia pastoris for high-level foreign gene expression." *Biotechnology* 12:181-184).

Plant systems may also be used for expression of perturbagens. Transcription of sequences encoding perturbagens may be driven by viral promoters, e.g. the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1991) "Deletion analysis of the 5' untranslated leader sequence of tobacco mosaic virus RNA." *J Virology* 65:1619-22). Alternatively, plant promoters such as that of the small subunit of RUBISCO or heat shock promoters may be used. (see, for example, Coruzzi, G. et al. (1984) "Tissue-specific and light-regulated expression of a pea nuclear gene encoding the small subunit of ribulose-1,5-bisphosphate." *EMBO J.* 3:1671-80; Broglie, R. et al. (1984) "Light-regulated expression of a pea ribulose-1,5-bisphosphate carboxylase small subunit gene in transformed plant cells." *Science* 24:838-843).

In each case described above, the selected construct can be introduced into the selected host cell by direct DNA transformation or pathogen-mediated transfection. The terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Preferred technologies for introducing perturbagens into mammalian cells include, but are not limited to, retroviral infection as well as transformation by EBV or similar episomally-maintained viral vectors (Makrides, S.C. (1999) "Components of vectors for gene transfer and expression in mammalian cells." *Protein Expr Purif* 17(2):183-202). Other suitable methods for transforming or transfecting host cells can be found in Maniatis, T. et al. ("Molecular Cloning: A Laboratory Manual." Cold Spring Harbor Laboratory Press) and other standard laboratory manuals.

For long term production of recombinant proteins in mammalian systems, stable expression of perturbagens in cell lines is preferred. For example, sequences encoding perturbagens can be transformed or introduced into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Alternatively, cells can be transfected using, for instance, retroviral, adenoviral, or adeno-associated viral agents as delivery systems for the perturbagen. For example, retroviral vectors (e.g. LRCX, Clontech) may be used to introduce and express perturbagens in a variety of mammalian cell cultures. Such vectors may rely on the virus' own 5' LTR as a means of driving perturbagen expression or may utilize alternative promoters/enhancers (e.g. those of CMV, RSV, and SV40, PIND) to regulate perturbagen expression levels.

In some instances, a preliminary selection is performed to verify that the host cells have been successfully transformed/transfected. Following the introduction of the vector, cells are allowed to grow in enriched media, and are then switched to selective media. The selectable marker confers resistance to the selective agent, and thus, only those cells that successfully express the introduced sequences survive in the selective media. Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in tk- or apr- cells, respectively (see e.g. Wigler, M. et al. (1977) "Transfer of purified herpes virus thymidine kinase gene to cultured mouse cells." *Cell* 11:223-32; Lowy, I. et al. (1980) "Isolation of transforming DNA: cloning the hamster aprt gene." *Cell* 22:817-23). Also antimetabolite, antibiotic, or herbicide resistance can be used as the basis for

selection. For example, *dhfr* confers resistance to methotrexate,; *neo* confers resistance to the aminoglycosides, neomycin and G-418, and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (see Wigler, M. et al. (1980) "Transformation of mammalian cells with an amplifiable dominant-acting gene." *PNAS* 77:3567-70; Colbere-Garapin, F. et al. (1981) "A new dominant hybrid selective marker for higher eukaryotic cells." *J. Mol. Biol.* 150:1-14). Additional selectable genes have been described, e.g. *trpB* and *hisD*, which alter cellular requirements for metabolites. Visible markers, e.g. anthocyanins, green, red or blue fluorescent proteins (Clontech), B glucuronidase and its substrate B glucuronide, or luciferase and its substrate luciferin, may also be used. Resistant clones containing stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Hose cells transformed/transfected with nucleotide sequence encoding for the perturbagen of interest may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. For example, the protein produced by a transformed transfected cell may be secreted when the selected expression vector incorporates signal sequences that direct secretion of the perturbagen through a prokaryotic or eukaryotic cell membrane.

Signal sequences also may be selected so as to direct the perturbagen to a particular intra-cellular compartment (Bradshaw, R.A. (1989) "Protein translocation and turnover in eukaryotic cells." *Trends Biochem Sci* 14(7):276-9). Perturbagen sequences may be isolated or purified from recombinant cell culture by methods heretofore employed for other proteins, e.g. native or reducing SDS gel electrophoresis, salt precipitation, isoelectric focusing, immobilized pH gradient electrophoresis, solvent fractionation, chromatography such as ion exchange, gel filtration, immunoaffinity and ligand affinity.

H. Host Cells

Host cell lines for use in the methodology described herein typically embody such desirable traits as 1) short cell cycle (i.e. 20-36 hr. doubling time), 2) amenability to high throughput procedures (e.g. FACS) without undue loss of membrane integrity or viability, 3) susceptibility to standard techniques designed to introduce various forms of foreign DNA. In addition, in order to use floater populations as a method of identifying and enriching for dead and/or dying cells in a negative selections, cell lines preferably display two additional features: (1) in a stable untreated cell population, the greater majority of cells are adherent to the solid support (e.g. plastic, gelatin) and the background rate of floater cells

is relatively low (<1%), (2) in an untreated or treated cell population (i.e. one exposed to putative cytotoxic agents and optionally, secondary agents), a high percentage of the floater cells correlate with the dead and/or dying cell population. One non-limiting example of a satisfactory and acceptable cell line is the colorectal adenocarcinoma cell line, HT29. HT29 (ATCC# HTB-38), is highly susceptible to retroviral infection and other methods of introducing foreign genetic materials and can express/maintain said materials for long periods of time using a variety of selectable markers common to the field (e.g. neomycin, puromycin). In addition, previous studies have shown that HT29 cultures transduced with control retroviral vectors exhibit low levels of floaters (<1%) and that a sizeable percentage of these cells (~40%) can be shown to be dead and/or dying (see USSN 09/504,132). Thus, HT29 serves as an acceptable cell line for isolating performing negative selection procedures.

In addition, it is understood that there are many other suitable host cell lines may be used in these studies. These lines include, but are not limited to, i) SW620 colon cancer cells (colorectal adenocarcinoma, ATCC # CCL-227), ii) the metastatic mammary epithelial cell tumor T-47D (ATCC # HTB-133), iii) Line 96C, and HMEC (Human Umbilical Vein Endothelial Cells,) line transformed with SV40 Lg T Ag, hTERT, and V12 H-Ras, iv) un-transformed, primary cell lines such as HuVECs and v) HMEC (Clonetics) and HCT15 (a colorectal adenocarcinoma, ATCC # CCL225). In some instances, these secondary cell lines may be used individually or in combination with other agents to study the specificity and application of perturbagens. For example, perturbagens isolated in HT29 cells may be introduced into other tumor cell lines (or primary cells) to determine whether the action of the perturbagen is specific to the cell line and genetic background in which it was isolated. Alternatively, perturbagens may be introduced into cells and studied in combination with any of several agents (e.g. camptothecin) to study whether the effects of a perturbagen can be augmented or sensitized by said secondary agents.

I. Screening for Biological Activity

The phenotypic assay (otherwise referred to as a "Floater Assay", see USSN 09/504,132) described herein selects for perturbagens that induce cell death. The procedures used to screen libraries for such perturbagens include: 1) introducing perturbagen encoding sequences (libraries) into the cell line of choice (e.g. HT29), 2) growing said cells under the appropriate conditions necessary to identify perturbagens that induce cell death, 3) collecting and/or screening said dead and/or dying cells by centrifugation, FACS or alternative high-throughput methods that provide efficient segregation of cells with the appropriate

phenotype, 4) re-isolating perturbagen encoding sequences from sorted cell populations by various techniques (e.g. PCR) and constructing new, sublibraries (e.g. retroviral sublibraries) from the PCR product; 5) enriching for perturbagens by recycling said sequences through the floater assay; and optionally 6) performing secondary assays to test specificity and scope of the agent.

In order to perform "floater assays" to identify sequences that encode cytotoxic agents (i.e., agents that stimulate relatively immediate death of individual cells, or agents that prevent cell growth or proliferation, thus gradually leading to the death of a cell population), libraries of sequences encoding putative cytotoxic/cytostatic agents are constructed and then introduced into the selected cell lines. In some instance, no further manipulation of the cultures is required. In other protocols, it may be possible to look for perturbagens that induce cell death under specific conditions. For example, it may be possible to identify perturbagens that only kill cells in the presence of a particular sensitizing agent (e.g. Camptothesin or Cis platinin) or under conditions where media supplements (e.g. insulin, or serum) are removed.

Libraries may take several forms including natural libraries (e.g. those constructed from cDNA and/or genomic DNA) or synthetic libraries. In most cases, libraries are introduced into the cell type of choice using common, state of the art, retroviral technology. While this is the preferred methodology for introduction of perturbagen encoding sequences into the cell type of choice, alternative methods that lead to transient expression of perturbagens, may also be used (e.g. REF).

To identify cells that contain cytotoxic sequences, floaters (e.g. cells that release from the solid support) are collected. This may be accomplished by pooling the media overlying the perturbagen infected cell cultures and centrifuging the samples at low speed to retrieve the floater population. Alternatively, non-adherent cells contained in the media can be stained with (for instance) propidium iodide, Apo2.7, or annexin, and then sorted by FACS to isolate all dead and/or dying cells.

Several methods may be used to retrieve the perturbagen sequences from cells that have been sorted. For instance, perturbagen-encoding sequences may be recovered by PCR (see, for example, Schott, B. (1997) "Efficient recovery and regeneration of integrated retroviruses." *Nucleic Acids Res.* 25(14):2940-2). In the case where perturbagen sequences are stably inserted into the host genome by retroviral infection, genomic DNA (derived from, for instance, cells taken from the FACS sorting procedures) is used as the template for PCR amplification. Complex mixtures with diversities of greater than 50,000 can be amplified

efficiently using oligonucleotide primers that flank the perturbagen encoding sequence. These sequences can subsequently be ligated into an appropriate retroviral vector, and re-introduced into a fresh population of e.g., HT29 cells for additional rounds of screening and enrichment.

5 Lastly, it is important to note that various methods and instrumentation familiar to those who are skilled in the art are used to construct, screen and test perturbagen libraries. The media, supplements, and reagents used in culturing, packaging, and maintenance of HT29 cells and HS293gp packaging cell lines can be purchased from a variety of commercial sources (Life Technologies, Clonetics, Cocalico Biologicals Inc., 10 ATCC). It should be noted that although a particular set of procedures and media formulations are used in the work described herein, alternatives can be substituted with little or no effect. For instance, in most cases, retroviral packaging was accomplished using $CaCl_2$. Though this is the preferred method of introducing retroviral vectors into 293gp packaging cells, alternative procedures such as LipofectAmine, may be applied. Molecular techniques 15 used in procedures such as genomic DNA isolation, PCR amplification, DNA endonuclease digestion, ligation, cloning, and sequencing, utilize common reagents that are supplied commercially (see, for example, Qiagen, New England BioLabs, Stratagene). When applicable, fluorescent activated cell sorting and analysis is performed on a Coulter EPICS Elite Cell Sorter using EXPO software. Again, alternative reagents and equipment, such as 20 MoFlo^R High-Speed Cell Sorter (Cytomation), are compatible with these procedures and may be substituted with little or no effect.

J. Cellular targets

In other embodiments, the invention encompasses the polypeptide, ribonucleotide, or polynucleotide sequence of the target (or fragment of each target) that is 25 identified with each perturbagen agent, as well as the gene encoding each target and relevant fragments of said gene. In this context, a "target" is any cellular component that is directly acted upon by the perturbagen that leads to and/or induces the phenotypic change.

Targets of specific perturbagens may be identified by several means. For instance, peptide perturbagens can be modified with homo- or hetero-bifunctional coupling 30 reagents and targets can be identified by chemical cross-linking techniques (see, for example, Tzeng, M.C. et al. (1995) "Binding proteins on synaptic membreanes for crotoxin and taipoxin, two phospholipases A2 with neurotoxicity." *Toxicon.* 33(4):451-7; Cochet, C. et al. (1998) "Demonstration of epidermal growth factor-induced receptor dimerization in living

cells using a chemical covalent cross-linking agent." *J Biol Chem.* 263(7):3290-5). Alternatively, one may use various techniques in column affinity chromatography or immunoprecipitation as a method of isolating and identifying target molecules (see, for example, Hentz, N.G. and Daunert, S. (1996) "Bifunctional fusion proteins of calmodulin and protein A as affinity ligands in protein purification and in the study of protein-protein interactions." *Anal Chem.* 68(22):3939-44). A preferred method involves application of a variation of the standard two-hybrid technology. See, e.g., U.S.S.N. 09/193,759 and WO 00/29565 "Methods for validating polypeptide targets that correlate to cellular phenotypes", the entire disclosures of which are incorporated by reference herein. Generally stated, the two-hybrid procedure is a quasi-genetic approach to detecting binding events. This assay often is performed in yeast cells (although it can be adapted for use in mammalian and bacterial cells), and relies upon constructing two vectors; the first having an interaction probe or "bait" (that in this case, is the perturbagen) that typically is fused to a DNA binding domain ("BD") moiety, and a second vector having an interaction target or "prey" (e.g. a cDNA library) that typically is fused to a DNA transcriptional moiety (the "activation domain" or "AD"). In an optimal setting, neither of the two fusion proteins can, individually, induce transcription of the reporter gene. Yet when the bait and prey interact, the AD and BD moieties are brought into sufficient physical proximity to result in transcription of a reporter gene (e.g., the *Leu2* gene or lacZ gene) located downstream of the bound complex (see Figure 4). Prey/bait interactions are then detected by identifying yeast cells that are expressing the reporter gene - e.g. which express lacZ or are able to grow in the absence of leucine.

A variety of yeast host strains known in the art are suitable for use for identifying targets of individual perturbagens. One of ordinary skill will appreciate that a number of factors may be considered in selecting suitable host strains, including but not limited to, (1) whether the host cells can be mated to cells of opposite mating type (i.e., they are haploid), and (2) whether the host cells contain chromosomally integrated reporter constructs that can be used for selections or screens (e.g., *His3* and *LacZ*). Although mating can be desirable in some embodiments, it is not strictly necessary for purposes of practicing the present invention. For example, the mating procedures can be eliminated by introducing the bait and prey constructs into a single yeast cell, whereupon the screens can be performed on the haploid cell.

Generally, either *Gal4* strains or *LexA* host strains may be used with the appropriate reporter constructs. Representative examples include strains yVT69, yVT87,

yVT96, yVT97, yVT98 and yVT99, yVT100, yVT360. Additionally, those of ordinary skill will appreciate that the host strains used in the present invention may be modified in other ways known to the art in order to optimize assay performance. For example, it may be desirable to modify the strains so that they can contain alternative or additional reporter genes that respond to two-hybrid interactions.

The following host yeast strains are thus constructed to have the indicated characteristics:

YVT69: yVT69 (mat , ura3-52, his3-200, ade2-101, trp1-901, leu2-3, 112, gal4D, met⁻, gal80D, URA3::GAL1$_{UAS}$-GAL1$_{TATA}$-lacZ) was obtained from Clontech (Y187).

YVT87: yVT87 (Mat-α ura3-52,his3-200,trp1-901,LexA$_{op(x6)}$-LEU2-3,112) was obtained from Clontech (EGY48).

YVT96: The starting strain was YM4271 (Liu, J. et al., 1993) MATa, ura3-52 his3-200 ade2-101 ade5 lys2-801 leu2-3, 112 trp1-901 tyr1-501 gal4D gal80D ade5::hisG. YM4271 was converted to yVT96, MATa ura3-52 his3-200 ade2-101 ade5 lys2:GAL2-URA3 leu2-3, 112 trp1-901 tyr1-501 gal4D gal80D ade5::hisG by homologous recombination of Reporter 1 to the LYS2 locus. The integration is confirmed by PCR.

YVT97: The starting strain is YM4271 (Liu, J. et al., 1993) MATa, ura3-52 his3-200 ade2-101 ade5 lys2-801 leu2-3, 112 trp1-901 tyr1-501 gal4D gal80D ade5::hisG. YM4271 will be converted to yVT97, MATα ura3-52 his3::GAL1 or GAL7-HIS3 ade2-101 ade5 lys2-801 leu2-3, 112 trp1-901 tyr1-501 gal4D gal80D ade5::hisG by the steps of (a) converting from MATa to MATα via transient expression of the HO endonuclease, *Methods in Enzymology* Vol. 194:132-146 (1991) and (b) integrating either of Reporters 3 or 4 at the HIS3 locus via homologous recombination. The integration is confirmed by PCR.

YVT98: The starting strain was EGY48 (Estojak, J. Et al., 1995) MATα, ura3 his3 trp1 leu2::LexAop(x6)-LEU2. EGY48 was converted to strain yVT98 MATα ura 3 his3 trp1 leu2::lexApo(x6)-LEU2 lys2::lexAop(8x or 2x)-LacZ by homologous recombination of Reporter 6 into the LYS2 locus.

YVT99: The starting strain was EGY48 (Estojak, J. et al., 1995) MATα, ura3 his3 trp1 leu2::LexAop(x6)-LEU2. EGY48 was converted to strain yVT99 MATa ura3 his3 trp1 leu2::lexAop(x6)-LEU2 lys2::lexAop(8x or 2x)-URA3 by homologous recombination of Reporter 2 into the LYS2 locus and by switching the mating type from MATα to MATa via transient expression of the HO endonuclease.

YVT100: The starting strain was YM4271 (Liu, J. et al., 1993) MATa, ura3-52 his3-200 ade2-101 ade5 lys2-801 leu2-3, 112 trp1-901 tyr1-501 gal4D gal80D ade5::hisG. YM4271 was converted to yVT100, MATa ura3-52 his3-200 ade2-101 ade5 lys2::lexAop(8x or 2x)-URA3 leu2-3, 112 trp1-901 tyr-501 gal4D gal80D ade5::hisG by homologous
5 recombination of Reporter 2 to the LYS2 locus. The integration was confirmed by PCR.

YVT360: yVT360 (mat a, trp1-901, leu2-3, 112, ura3-52, his3-200, gal4D, gal80D, LYS2::GAL1$_{UAS}$-GAL1$_{TATA}$-HIS3, GAL2$_{UAS}$-GAL2$_{TATA}$-ADE2, URA3:MEL1$_{UAS}$-MEL1$_{TATA}$-lacZ) was obtained from Clontech (AH109).

Exemplary yeast-reporter strains are constructed using a variety of standard
10 techniques. Many of the starting yeast strains already carry multiple mutations that lead to an auxotropic phenotype (e.g. ura3-52, ade2-101). When necessary, reporter constructs can be integrated into the genome of the appropriate strain by homologous recombination. Successful integration can be confirmed by PCR. Alternatively, reporters may by maintained in the cells episomally.

15 The yeast two-hybrid reporter gene typically is fused to an upstream promoter region that is recognized by the BD, and is selected to provide a marker that facilitates screening. Examples include the *lacZ* gene fused to the *Gal1* promoter region and the *His3* yeast gene fused to *Gal1* promoter region. A variety of yeast two-hybrid reporter constructs are suitable for use in the present invention. One of ordinary skill will appreciate that a
20 number of factors may be considered in selecting suitable reporters, including whether (1) the reporter construct provides a rigorous selection (i.e., yeast cells die in the absence of a protein-protein or peptide-protein interaction between the bait and prey sequences), and/or (2) the reporter construct provides a convenient screen (e.g., the cells turn color when they harbor bait and prey sequences that interact). Examples of desirable reporters include (1) the
25 *Ura3* gene, which confers growth in the absence of uracil and death in the presence of 5-fluoroorotic acid (5-FOA); (2) the *His3* gene, which permits growth in the absence of histidine; (3) the *LacZ* gene, which is monitored by a colorimetric assay in the presence/absence of beta-galactosidase substrates (e.g. X-gal); (4) the *Leu2* gene, which confers growth in the absence of leucine; and (5) the *Lys2* gene, which confers growth in the
30 absence of lysine or, in the alternative, death in the presence of α-aminoadipic acid. These reporter genes may be placed under the transcriptional control of any one of a number of suitable cis-regulatory elements, including for example the *Gal2* promoter, the *Gal1* promoter, the *Gal7* promoter, or the *LexA* operator sequences.

In other embodiments, perturbagens-induced phenotypes may be the result of RNA-polypeptide or polypeptide-DNA interactions. In cases such as these, variations of the original two-hybrid theme may be applied to identify the target of the phenotypic probe. (See, for example, Li, J.J. and Herskowitz, I. (1993). Isolation of Orc6, a Component of the Yeast Origin Recognition Complex by a One-Hybrid System. *Science,* 262:1870-1874; Svinarchuk, F. et al. (1997) "Recruitment of transcription factors to the target site by triplex-forming oligonucleotides." *NAR* 25:3459-3464; Sengupta D.J. et al. (1999) "Identification of RNAs that bind to a specific protein using the yeast three-hybrid system." *RNA* 5:596-601; Harada, K. et al. (1996) "Selection of RNA-binding peptides in vivo." *Nature* 14;380(6570):175-9; SenGupta D.J. et al. (1996) "A three-hybrid system to detect RNA protein interactions in vivo." *PNAS* 93:8496-8501).

Target sequences or fragments thereof can vary greatly in size. Some target fragments can be as small as ten amino acids in length. Alternatively, target sequences can be greater than 10 amino acids but less than thirty amino acids in length. Still other targets can be greater than thirty amino acids in length but shorter than 60 amino acids in length. Still other targets are cellular proteins or subunits or domains therein of approximately 60 amino acids in length. Still other targets are cellular proteins or subunits or domains therein of more than 60 amino acids in length. In addition, for reasons described previously, the sequences encoding targets can vary greatly due to allelic variation, duplications and loosely related gene family members. That said, the invention also encompasses variants of said targets. A preferred target variant is one which has at least about 80%, alternatively at least about 90%, and in another alternative at least about 95% amino acid sequence identity to the original target amino acid sequence and which contains at least one functional or structural characteristic of the original target.

K. Screening Assays

The agents of the invention can be used to screen for drugs or compounds (small molecules) that mimic, or modulate the activity or expression of said phenotypic probes. The term "small molecule," as used herein, refers to a chemical compound, for instance a peptide or oligonucleotide that may optionally be derivatized, or any other low molecular weight organic compound, either natural or synthetic. Such small molecules may be a therapeutically deliverable substance or may be further derivatized to facilitate delivery. Like the perturbagen itself, such compounds may be used to treat disorders characterized by insufficient or excessive production of a target which has decreased or aberrant activity

compared to the wild type entity. Thus, the invention provides a method for identifying modulators, i.e. candidate or test compounds or agents (e.g. peptidomimetics, small molecules or other drugs) that bind to the agent or its target, and have a stimulatory or inhibitory effect on the pathway(s) affected by said agent.

5 In one embodiment, the invention provides libraries of test compounds. The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries, spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the one-bead one-compound library method; and synthetic library 10 methods using affinity chromatography selection. The biological library approach is exemplified by peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S. (1997) "Application of combinatorial library methods in cancer research and drug discovery." *Anticancer Drug Des.* 12:145).

15 Methods for the synthesis of molecular libraries can be found in the art, for example, in (i) De Witt, S.H. et al. (1993) "Diversomers: an approach to nonpeptide, nonoligomeric chemical diversity." *PNAS* 90:6909, (ii) Erb, E. et al. (1994) "Recursive deconvolution of combinatorial chemical libraries." *PNAS* 91:11422, (iii) Zuckermann, R.N. et al. (1994) "Discovery of nanomolar ligands for 7-transmembrane G-protein-coupled 20 receptors from a divers N-(substituted)glycine peptoid library." *J. Med Chem.* 37:2678 and (iv) Cho, C.Y. et al. (1993) "An unnatural biopolymer." *Science* 261:1303. Libraries of compounds may be presented in i) solution (e.g. Houghten, R.A. (1992) "The use of synthetic peptide combinatorial libraries for the identification of bioactive peptides." *BioTechniques* 13:412) ii) on beads (Lam, K.S. (1991) "A new type of synthetic peptide library for 25 identifying ligand-binding activity." *Nature* 354:82), iii) chips (Fodor, S.P. (1993) "Multiplexed biochemical assays with biological chips." *Nature* 364:555), iv) bacteria (U.S. Patent # 5,223,409), v) spores (Patent Nos 5,571,698, 5,403,484, and 5,223,409), vi) plasmids (Cull, M.G. et al. (1992) "Screening for receptor ligands using large libraries of peptides linked to the C terminus of the lac repressor." *PNAS* 89:1865) or vii) phage (Scott, J.K. and 30 Smith, G.P. (1990) "Searching for peptide ligands with an epitope library." *Science* 249:386).

There are several methods for identifying small molecule compounds that mimic the action of the phenotypic probes. In one example, small molecule mimetics are identified using displacement assays. Such assays can be based upon a variety of

technologies including, but not limited to, i) ELISAs (see, for example, Rice, J.W. et al. (1996) "Development of a high volume screen to identify inhibitors of endothelial cell activation." *Anal Biochem* 241(2):254-9), ii) scintillation proximity assays (see, for example, Lerner, C.G. and Saiki, A.Y.C. (1996) "Scintillation proximity assay for human DNA topoisomerase I using recombinant biotinyl-fusion protein produced in baculovirus-infected insect cells." *Anal Biochem* 240(2):185-96) or iii) time-resolved fluorescence resonance energy transfer-based technology (see, for example, Fernandes, P.B. (1998) "Technological advances in high-throughput screening." *Curr Opin Chem Biol* 2(5):597-603; Hemmilä, "Time-resolved fluorometry - advantages and potentials in high throughput screening assays." *"High Throughput Screening"*, J. Devlin (ed.). Marcel Dekker Inc, New York, pp. 361-76 (1997)). Two non-limiting examples of such assays, one homogeneous, LANCE™ (Stenroos, K. et al. (1997) "Homogeneous time resolved fluorescence energy transfer assay (LANCE) for the determination of IL-2-IL-2 receptor interaction." Abstract of Papers presented at the 3rd Annual Conference of the Society for Biomolecular Screening, Sep., California), and one heterogeneous, DELFIA™ (MacGregor, I. et al. (1999) "Application of a time-resolved fluoroimmunoassay for the analysis of normal prion protein in human blood and its components." *Vox Sang* 77(2):88-96; Jensen, P.E. et al. (1998) "A europium fluoroimmunoassay for measuring peptide binding to MHC class I molecules." *J. Immunol. Methods* 215:71-80; Takeuchi, T. et al. (1995) "Nonisotopic receptor assay for benzodiazepine drugs using time-resolved fluorometry." *Anal. Chem.* 67:2655-8.) are described as follows:

1. LanceTM: Homogeneous Assay

To identify small molecules capable of disrupting the interaction between the perturbagens and its target, assays are designed to utilize the LANCE™ technology (commercially available from E. G. & G. Wallac.). LANCE™ is a homogeneous assay that is performed in solution and requires no wash steps to separate bound and unbound label. Briefly, the target is produced in large quantities and labeled with a lanthanide chelate (i.e. a fluorescent donor such as a Europium, (Eu) or Terbium (Tb) chelate). Concomitantly, the perturbagens is labeled with one of several fluorescent "acceptor" moieties that can be excited by the emissions of the donor molecule (e.g. allophycocyanin (APC) or rhodamine Rh, respectively). Most preferably, 1) the modifications of either the perturbagen or the target is not detrimental to the interaction between the two interacting molecules being studied and 2) the distance separating the donor and acceptor moieties when the perturbagen and the target are associated, is sufficiently close to permit FRET (typically 30-100

Angstroms). As an alternative to direct labeling of the perturbagen, monoclonal antibodies directed against the perturbagen can be labeled with Eu, thus allowing small molecule displacement assays to take place via indirect labeling procedures.

To identify small molecules capable of disrupting the interaction between the perturbagen and its target, the two labeled components are alliquoted into wells (1536 well format) at previously set, optimized conditions that will ensure 50% binding (see Figure 5). Subsequently, each well is then exposed to one or more members of a large chemical combinatorial library and time-resolved measurements are taken using a Wallac 1420 Victor multilabel counter or equivalent fluoremeter. In wells that contain a small molecule that interferes with the interaction between the perturbagen and its target, the distance separating the donor and acceptor molecules is increased. As a result of this dissociation or displacement, the ability of the Eu emissions to excite the acceptor is compromised and the total fluorescence emitted by the acceptor is decreased.

2. DELFIA™: Heterogeneous Assay

Several variations of a heterogeneous assay (DELFIA™) using an immobilized substrate can be used as an alternative to LANCE™. In one non-limiting example, the target is immobilized to a solid support using a monoclonal antibody that has been labeled with Eu (see Figure 6). Subsequent addition and binding of a rhodamine labeled perturbagen in the presence or absence of a candidate small organic displacement molecule is followed by several wash steps to remove unbound material. TR-FRET is then performed by exciting Eu and measuring the levels of Rh emissions. As an alternative to this procedure, the target is immobilized to the solid support using an unlabeled monoclonal antibody. Subsequently, an Eu-labeled perturbagen (+/- a candidate small organic displacement molecule) is added to each well and allowed to equilibrate, followed by a washing procedure to eliminate unbound Eu-labeled material. Once the well has been cleared of all unbound material, the bound Eu-perturbagen molecules are released and excited in the presence of commercially available enhancement solution (DELFIA™ Enhancement Solutions, Wallac). By comparing the levels of emissions in wells that contain members of the molecule library with standardized controls, small molecules that disrupt the interaction between the perturbagen and its target are identified.

L. Therapeutic Uses

Natural and synthetic chemotherapeutic derivatives have proven valuable in the treatment of a variety of forms of disease. For that reason, in one embodiment,

perturbagens, fragments or derivatives of a perturbagen, small molecule mimetics of a perturbagen, sequences encoding perturbagens, sequences that can hybridize to perturbagen encoding sequences, targets of the perturbagen, or agents that bind said target (e.g. antibodies) or portions thereof, may be utilized to treat or prevent a disorder that has
5 previously shown sensitivity to treatment with chemotherapeutics and/or radiation therapy. Thus, for example, polypeptides or RNA molecules described herein can be used i) modulate cellular proliferation, ii) module cellular differentiation, iii) induce or modulate necrotic or apoptotic processes, or iv) sensitize cells to secondary compounds that induce either i), ii), or iii) by direct application of said agent. Examples of such disorders that may be aided by such
10 agents include, but are not limited to cancers of the i) ovary, ii) liver, iii) endometrium, iv) stomach, colon and/or rectum, v) prostrate, vi) uterus, vii) esophagus, viii) kidney, ix) thyroid, x) stomach, xi) brain, xii) skin and xiii) breast.

Ailments such as those described previously can be treated with the perturbagen directly or indirectly. Thus, either a purified form of the perturbagen can be
15 administered to the patient or a vector capable of expressing a perturbagen or a fragment or derivative thereof may be administered to a subject to treat or prevent a disease. Expression vectors including, but not limited to, those derived from retroviruses, adenoviruses, adeno-associates viruses, or herpes or vaccinia viruses or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population (see,
20 for example, Carter, P.J. and Samulski, R.J. (2000) "Adeno-associated viral vectors as gene delivery vehicles." *Int J Mol Med.* 6(1):17-27; Paul, G. et al. (2000) "Progress with retroviral gene vectors." *Rev Med Virol.* 10(3):185-202; Wu, N. and Ataai, M.M. (2000) "Production of viral vectors for gene therapy application." *Curr Opin Biotechnol.* 11(2):205-8).

In a further embodiment, a pharmaceutical composition comprising a
25 substantially purified perturbagen, or a fragment thereof, or a small molecule mimetic, optionally in conjunction with a suitable pharmaceutical carrier, may be administered to a subject to treat or prevent any of the previously mentioned disorders. As used herein, the language "pharmaceutical carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, and the like, compatible with
30 pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated.

Pharmaceutical compositions of the invention are formulated to be compatible with intended routes of delivery. Examples of routes of administration include parenteral e.g.

intravenous, intradermal, subcutaneous, oral, inhalation, transdermal, topical, transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent, such as water for injection, saline solution, fixed oils, polyethylene, glycols, glycerine, propylene glycol, or other synthetic solvents, antibacterial agents such as benzyl alcohol or methyl parabens, antioxidants such as ascorbic acid or sodium bisulfite, chelating agents such as ethylenediaminetetraacetic acid, buffers such as acetates, citrates, or phosphates and agents for the adjustments of tonicity such as sodium chloride or dextrose.

Pharmaceutical compositions suitable for injectable use include aqueous solutions (where water-soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water Cremophor EL™ (BASF; Parsippany, NJ) or phosphate buffered saline (PBS). In all cases the composition must be sterile and should be fluid to the extent that easy syringability exists. Oral compositions can also be prepared using any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth, or gelatin; an excipient such as starch or lactose, disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate, a glidant such as colloidal silicon dioxide, a sweetening agent such as sucrose or saccharin, or a flavoring agent such as peppermint or orange flavoring. For administration by inhalation, the compounds are delivered in the form of an aerosol spray from a pressurized container or dispenser that contains a suitable propellant. Systemic administration can also by transmucosal or transdermal means. For these methods of administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art and include, for example, bile salts and fusidic acid derivatives. Transmucosal administration can also be accomplished through the use of nasal sprays and suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled microencapsulated delivery system. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation

and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to specific cell surface epitopes) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

5 The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (U.S. Patent 5,328,470) or by stereotactic injection (see, for example, Chen, S.H. et al. (1994) "Gene therapy for brain tumors: regression of experimental gliomas by adenovirus-mediated gene transfer in vivo." *PNAS* 10 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g. retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery 15 system.

M. Diagnostic Uses

The polynucleotides, polypeptides, variants, targets and antibodies to any one of these molecules can, in addition to previously mentioned therapeutic applications, be used in one or more of the following methods: 1) detection assays (e.g. chromosomal mapping, 20 tissue typing, forensic biology), and 2) predictive medicine (e.g. diagnostic or prognostic assays, pharmacogenomics and monitoring clinical trials). Thus, for example, agents may be used to detect a specific mRNA or gene (e.g. in a biological sample) for a genetic lesion. Similarly, agents described herein may be applied to the field of predictive medicine in which diagnostic assays or prognostic assays, pharmacogenomics, and monitoring clinical trials are 25 used for predictive purposes to thereby treat an individual prophylactically.

Accordingly, one aspect of the present invention relates to diagnostic assays for determining expression of a polypeptide or nucleic acid of the invention and/or activity of said agent of the invention, in the context of a biological sample to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a 30 disorder, associated with aberrant expression or activity of a polypeptide or polynucleotide of the invention.

Alternatively, the invention provides methods for detecting expression of a nucleic acid or polypeptide of the invention or activity of a polypeptide or polynucleotide of

the invention in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (e.g. drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (e.g. the genotype of the individual examined to determine the ability of the individual to respond to a particular agent). Still another aspect of the invention pertains to monitoring the influence of agents (e.g. drugs or other compounds) on the expression or activity of a polypeptide or polynucleotide of the invention in clinical trials.

1. Detection Assays

Portions or fragments of the polynucleotide sequences of the invention can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to i) map their respective genes on a chromosome and, thus, locate gene regions associated with genetic diseases; ii) identify an individual from a minute biological sample (tissue typing); and iii) aid in forensic identification of biological samples.

a. Gene and Chromosome Mapping

Once the sequence (or portion of a sequence) of a gene has been isolated, this sequence can be used to identify the entire gene, analyze the gene for homology to other sequences (i.e., identify it as a member of a gene family such as EGF receptor family) and then map the location of the gene on a chromosome. Accordingly, nucleic acid molecules described herein or fragments thereof, can be used to map the location of the gene on a chromosome. The mapping of the sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease. Briefly, genes can be mapped to chromosomes by preparing PCR primers from the sequence of a gene of the invention. These primers can then be used for PCR screening of somatic cell hybrids containing individual chromosomes. Only those hybrids containing the human gene corresponding to the gene sequences will yield an amplified fragment (For review of this technique see D'Eustachio, P. and Ruddle, F.H. (1983) "Somatic cell genetics and gene families." *Science* 220:919-924). Alternative methods of mapping a gene to its chromosome include in situ hybridization (see, for example, Fan, Y.S. et al. (1990) "Mapping small DNA sequences by fluorescence in site hybridization directly on banded metaphase chromosomes." *PNAS* 87:6223-27), pre-screening with labeled flow sorted chromosomes (CITE), and pre-selection by hybridization to chromosome specific cDNA libraries. Furthermore, fluorescence in site hybridization (FISH) of a DNA sequence to a metaphase chromosome spread can further be used to provide a precise chromosomal location in one step (see

"Human Chromosomes: A Manual of Basic Techniques", Pergamon Press, New York, 1988). Lastly, with the completion (in the not-to-distant future) of the sequencing of the human genome, chromosome mapping will very quickly switch from elaborate, hands-on methods of mapping genes, to simple database searches.

5 Once the sequence (or portion of a sequence) of a gene has been isolated, these agents can be used to assess the intactness or functionality of a particular gene. Comparison of affected and unaffected individuals can begin with looking for structural alterations in the chromosomes such as deletions, inversions, or translocations that are based on that DNA sequence. Once this is accomplished, the physical position of the sequence on 10 the chromosome can be correlated with genetic data map. (Such data are found, for example, in McKusick, V. "Mendialian Inheritance in Man" available on-line through John Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in e.g. Egeland, J.A. et al. (1987) "Bipolar affective 15 disorders linked to DNA markers on chromosome 11." *Nature,* 325:783-787). Alternatively, polynucleotide sequences can be used as probes in Southern Blot analysis to identify alterations in the organization of the gene of interest and surrounding regions. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms. If a specific 20 mutation is observed in some or all individuals affected by a particular disease, but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease.

b. Tissue Typing

The nucleic acid sequences of the present invention can also be used to 25 identify individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. The sequences of the present invention are useful as additional DNA 30 markers for RFLP mapping (described in US Patent 5,272,057).

Furthermore the sequences of the present invention can be used to determine the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the nucleic acid sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the individual's DNA and subsequently sequence it. Panels of

corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic variation. The sequences of the present invention can be used to obtain such identification sequences from individuals and from tissue. The nucleic acid sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the non-coding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per 500 bases. Thus, each of the sequences described herein may be, to some degree, used as a standard against which DNA from an individual can be compared for identification purposes.

c. Forensic Biology

In addition, the sequences described herein can be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example a perpetrator of a crime. To make such an identification, PCR-based technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, (e.g. hair, skin, or body fluids). The amplified sequence can then be compared to a standard thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents (e.g. PCR primers) targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that is unique to a particular individual. The nucleic acid sequences described herein can further be used to provide polynucleotide reagents e.g. labeled or labelable probes, which can be used in, for example, an in situ hybridization technique, to identify a specific tissue. This technique can be exceedingly useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such probes can be used to identify tissue by species and/or organ type.

N. Predictive Medicine

Portions or fragments of the polynucleotide sequences of the invention can be used for predictive purposes to thereby treat an individual prophylactically.

1. Diagnostic/Prognostic Assays

One method of detecting the presence or absence of a polypeptide or nucleic acid in a biological sample is to expose that sample to an agent that recognizes the entity in question. A preferred agent for detecting mRNA or genomic DNA is labeled nucleic acid probe capable of hybridizing to the sequence one is attempting to detect (for instance, the sequence of the invention). The nucleic acid probe can be, for example, a full length cDNA, or a portion thereof such as an oligonucleotides of at least 15, 30, 50, 100, 250, or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to a mRNA or genomic DNA encoding the invention. The term "labeled" in this context refers to modifications in said sequences including, but not limited to, biotin labeling that can then be detected with a fluorescently labeled streptavidin, or ^{32}P labeling.

A preferred agent for detecting a polypeptide of the invention is an antibody or peptide capable of binding to the invention, preferably an antibody with a detectable label. Antibodies can be polyclonal or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g. a Fab or $F(ab)_2$) can be used. The term "labeled" in this context refers to direct labeling of the probe or antibody by coupling (i.e. physical linking) a detectable substance to the probe or antibody, such as a fluorescent labeled moiety or biotin.

The detection methods of the invention can be used to detect mRNA, protein, or genomic DNA in a biological sample in vitro as well as in vivo. For example, in vitro techniques for detection of mRNA include (but are not limited to) Northern Blot hybridization and in situ hybridizations. In vitro techniques for detection of a polypeptide of the invention include enzyme linked immunosorbent assays (ELISA's), Western blots, immunoprecipitations, and immunofluorescence.

The invention also encompasses kits for detecting the presence of a polypeptide or nucleic acid of the invention in a biological sample. Such kits can be used to determine if a subject is suffering from or is at increased risk of developing a disorder associate with aberrant expression of a polypeptide or polynucleotide of the invention. For instance, the kit can comprise a labeled compound or agent (as well as all the necessary supplementary agents needed for signal detection e.g. buffers, substrates, etc...) capable of detecting the polypeptide, or mRNA in the sample (e.g. an antibody which binds the polypeptide or a oligonucleotide probe that binds to DNA or mRNA encoding the polypeptide).

The methods of the invention can also be used to detect genetic lesions or mutations in a gene of the invention, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized by aberrant expression or activity of an agent of the

invention. In preferred embodiments, the methods include detecting the presence or absence of a genetic lesion or mutation characterized by at least one alteration affecting the integrity of the agent of the invention. For example, such genetic lesions or mutations can be detected by ascertaining the existence of at least one of: 1) a deletion of one or more nucleotides from a gene; 2) an addition of one or more nucleotides to a gene; 3) a substitution of one or more nucleotides of the gene; 4) a chromosomal rearrangement of the gene; 5) an alteration in the level of a messenger RNA transcript of the gene; 6) an aberrant modification of the gene, such as of the methylation pattern of the genomic DNA; 7) the presence of a non-wild type splicing patters of a messenger RNA; 8) a non-wild type level of the protein encoded by the gene; 9) an allelic loss of the gene; and 10) an inappropriate post translational modification of the protein encoded by the gene. Many techniques can be used to detect lesions such as those described above. For instance, mutations in a selected gene from a sample can be identified by alterations in restriction enzyme cleavage patterns. In this procedure, sample and control DNA is isolated, digested with one or more restriction endonucleases, and fragment length sizes (determined by gel electrophoresis) are compared. Observable differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Additional techniques that can be applied to detecting mutations include, but are not limited to, detection based on direct sequencing, PCR-based detection of deletions, inversions, or translocations, detection based on mismatch cleavage reactions (Myers, R.M. et al. (1985) "Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA:DNA duplexes." *Science* 230:1242), and detection based on altered eletrophoretic mobility (e.g. SSCP, see for example, Orita, M. et al. (1989) "Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms." *PNAS* 86:2766).

2. Pharmacogenetics

Pharmacogenetics deals with clinically significant hereditary variation in the response to drugs due to altered drug disposition and altered action in affected persons (see Linder, M.W. et al. (1997) "Pharmacogenetics: a laboratory tool for optimizing therapeutic efficiency." *Clin Chem.* 43(2):254-266). In general, two types of pharmacogenetic conditions can be differentiated. There are genetic conditions transmitted as a single factor altering the way drugs act on the body, referred to as "altered drug action". Alternatively, there are genetic conditions transmitted as single factors altering the way the body acts on drugs (referred to as "altered drug metabolism"). These two conditions can occur either as rare defects, or as polymorphisms. For example, glucose-6-phosphate dehydrogenase

deficiency is a common inherited enyzmopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (e.g. anti-malarials, sulfonamides etc.).

The activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (e.g. N-acetyltransferase 2 (NAT2) and cytochrome P450 enzymes (CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM which all lead to the absence of functional CYP2D6. Poor metabolizers of this sort quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, a PM will show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. At the other extreme are the so-called ultra rapid metabolizer who do not respond to standard doses. Recently, the molecular basis of ultra rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, in the context of pharmacogenetics, an agent of the invention can be used to determine or select appropriate agents for therapeutic prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individuals drug responsiveness phenotype.

3. Monitoring of Effects During Clinical Trials

Monitoring the influence of agents that effect the expression or activity of a polypeptide or polynucleotide of the invention can be applied in clinical trials. For example, the effectiveness of a drug directed toward a target identified by the invention and intended to treat a particular ailment, can be monitored in clinical trials of subjects exhibiting said ailment by monitoring the level of gene expression of the target, activity of the target, or levels of the target of the invention. Thus, in a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent by comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of the polypeptide or polynucleotide of the

invention in the pre-administration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level or activity of said target of the invention in the post administration samples, (v) comparing the level of said target of the invention in the post-administration sample with levels in the pre-administration samples, and (vi) altering the
5 administration of the agent to the subject accordingly.

EXAMPLES

The following examples are intended to further illustrate certain preferred embodiments of the invention, and are not limiting in nature.

Example 1
10 **Preparation of a cDNA Library**

To construct a retroviral vector that was appropriate for negative selections, the 3.8kB HindIII/ScaI band of pVT314 was ligated to the 1.9kB SSPI/PvuII band of pBluescript™ (Stratagene). The final product of this reaction (referred to as pVT340; see Figure 7) contained all the necessary components of a constitutive retroviral expression
15 vector including a Psi site for packaging, constitutive CMV driven perturbagen expression, a splice donor and acceptor site for obtaining high levels of library insert expression, and a multiple cloning site (MCS) linked to the 3' end of EGFP. As a source of perturbagens, a cDNA library obtained from human brain tissue was purchased from Origene Inc. (Catalogue #DHL 101, DHL 105, and DHL 106) and transformed into bacteria (DH10B, Gibco). Using
20 standard techniques, bacterial hosts carrying the libraries were then expanded in liquid media (LB plus ampicillin) and used to prepare large quantities of episomal (library) DNA (Maxiprep, Qiagen). The cDNA insert in each vector was then released by digestion with the appropriate restriction enzymes (EcoRI/XhoI) and the fragments measuring 0.4-2.8kB were gel purified and ligated (T4 Ligase, Boehringer Mannheim) into the compatible sites of the
25 pVT340 retroviral vector. As a result of these procedures, putative cytotoxic agents were expressed constitutively as fusions with the GFP scaffold.

EXAMPLE 2

Retroviral Packaging and Infection

Library constructs were packaged for retroviral transfection into the cell of
30 choice using LipofectAMINE. Specifically, on Day 1, 3 x 10^6 cells of the packaging cell line (293gp) were seeded into a T175 flask. On the second day, two tubes, one carrying 15ug of

library DNA + 10ug of envelope plasmid (pCMV-VSV.G-bpa) + 1.5 ml DMEM (serum free), and the second carrying 100ul of LipofectAMINE (Gibco BRL) + 1.5ml DMEM (serum free) were mixed and left at room temperature for 30 minutes. Subsequently, the two tubes were mixed together along with 17 ml of serum free DMEM. This cocktail is referred to as the "transfection mix." Previously plated 293gp cells were then gently washed with serum free media and exposed to 20 ml of the transfection mix for 4 hours at 37°C. Following this period, the transfection mix was removed and the cells were incubated with complete DMEM (10% serum) for a period of 72 hours at 37°C. On Day 4 or 5, the media (now referred to as "viral supernatant") overlying the 293gp cells was collected, filtered through a 0.45μ filter, and frozen down at -80°C.

As an alternative to the LipofectAMINE method of retroviral DNA packaging, a second protocol, referred to herein as the "$CaCl_2$ Method," can be used to package retroviral sequences. In this method, 5 x 10^6 cells of the packaging cell line (293gp) are seeded into a 15cm^2 flask on Day 1. On the following day, the media is replaced with 22.5 mls of modified DMEM. Subsequently, a single tube carrying 22.5μg of retroviral library DNA and 22.5μg of envelope expression plasmid (pCMV-VSV.G-bpa) is brought to 400μl with dH_2O, to which is added 100μl of $CaCl_2$ (2.5M) and 500μl of PBS (dropwise addition, 2x solution = 50mM, BES (N,N-bis(2-hydroxyethyl)-2-aminoethane-sulfonic acid), 280mM NaCl, 1.5mM Na_2HPO_4, pH 6.95). After allowing this retroviral mixture to sit at room temperature for 5-10 minutes, the mixture is added to the 293gp cells in a dropwise fashion, and the cells are then incubated at 37°C (3% CO_2) for 16-24 hours. The media is then replaced and the cells are allowed to incubate for an additional 48-72 hours at 37°C. At that time, the media containing the viral particles is then collected, filtered through a 0.45μ filter and frozen down at -80°C.

EXAMPLE 3

Floater Cell Assays in HT29 Colon Cancer Cells

In the first round of selection, twenty T175 flasks were seeded with 2.2 x 10^6 HT29 cells/flask in McCoy's 5A media (Gibco BRL) modified with 10% FBS. On Day 1, each flask was infected with a retroviral supernatant (4μg/ml polybrene, 50% volume) containing the cDNA library and on Day 2 the media was changed. Using the fluorescent properties of GFP as an indicator of infection levels, FACS analysis was performed on a sample of the population. In cycles 1-6, approximately 90% of the viable cells were observed

to be infected with the retroviral library. In cycle 7, cells were intentionally infected with low moieties of infection (MOI) to ensure single inserts in each cell. As a result, FACS analysis showed that only 5% of the cells contained a viral insert.

On Day 3 and 5, floater cells (totaling roughly 1% of the total population) from the original twenty flasks were divided into three separate subpools (I, II, and III), and readied for genomic DNA (gDNA) preparation using a QIAamp kit (Qiagen). Briefly, floater cells were centrifuged into a pellet (400 x g), washed in PBS, and lysed to release gDNA. This material was then gravity filtered over a QIAamp column designed to bind/retain gDNA. The column was then washed several times to remove protein and RNA contamination, and the gDNA was then eluted with dH_2O and treated. Genomic DNA samples were then ethanol precipitated, washed, and treated with RNAse A to eliminate any RNA contamination.

Floater cell gDNA was then subjected to PCR procedures to amplify the library sequences encoded therein. Briefly, the above gDNA aliquot was divided into 1 µg samples for use as templates for PCR using the oligonucleotides OVT 800 (5' GCCGCCGGGATCACTCTC) (SEQ ID NO: 1) and OVT 1211 (5' GCTAGCTTGC CAAACCTACAGGTGGGG) (SEQ ID NO: 2) (PCR conditions were: 95°C, 30 seconds; 95°C, 15 seconds; 63°C, 30 seconds; 72°C, 3 minutes, cycle to "Step 2" twenty four times; and 72°C, 5 minutes). The resulting PCR products were then purified using QIAquick (Qiagen), digested with EcoRI and XhoI, and then directionally ligated into the original retroviral vector (pVT340). This material was then transformed into electrocompetent bacterial cells (DH10B, Gibco BRL) and plated out on LB-amp plates. Each sublibrary was subsequently grown in liquid culture (LB + ampicillin), processed to yield plasmid DNA (Quiagen Maxi Prep), and repackaged in 293gp cells. The resulting viral supernatants were then reinfected into naive HT29 cells (1×10^6 HT29 cells per flask, three flasks per subpool, 50% supernatants) to begin the second round of negative selection. Cycles 2-6 used protocols and techniques similar to those described above. In cycle 7, both the numbers of cells plated and percent viral supernatant used in the infection of the cells were reduced (cycle 7: 2 T75 flasks per subpool, 6.5×10^5 cells per flask, Flask #1 = 10% supernatant, Flask #2 = 1% supernatant).

Results from seven consecutive cycles of floater selections are summarized as follows: i) both mock infected cells and pVT340 control vector cells consistently showed 1% (or less) floaters in the media and ii) all three subpools exhibited a steady increase in the percent floater population over the course of the cycling (see Figure 8). On average, the number of floaters grew to roughly 14% by the end of cycle 7. This data demonstrates

successful enrichment for perturbagen sequences that increase the frequency of dead and/or dying HT29 cells.

Thirteen hundred and forty seven clones derived from cycle 7 were sequenced and analyzed. The sequence sets revealed high enrichment for two types of Bid clones (BidA and BidB; see Figure 9). By cycle 7, over 20% of the individual clones in the sublibrary encoded one of the two in-frame fragments of the pro-apoptotic protein Bid (Wang et a., 1996). When re-tested, these clones produced a marked increase in floater percent compared with controls. Moreover, when the sequence of clones taken from multiple stages of the enrichment procedure were compared, it was found that one Bid clone (BidB) increased nearly 36-fold in abundance between cycles 6 and 7, thus demonstrating the validity of cycling as an enrichment procedures.

The predominance of Bid clones in the later stages of the enrichment procedures suggested the possibility that weaker clones may have been lost during the final selection cycles, perhaps excluded by the strong clones such as Bid. This disparity in representation should be exponential (a power of cycle number), and should be less pronounced in earlier cycles. Thus, the cycle 4 sublibrary, the first one in which significant increases in floater rate were observed, was chosen as a source of cytotoxic/cytostatic perturbagens.

Because of the low floater rate displayed by cells transduced with cycle 4 material, and the likely corresponding low frequency of cell-lethal clones, a thorough search for clones, both weakly penetrant ones and strong ones, necessitated development of a large-scale system with the capacity to analyze thousands of single sequences. To solve this throughput problem, an automated genetic analysis system for mammalian cells called Somata was developed (See U.S. provisional application Serial No. 60/305712, "Automated Assay Methodology," the contents of which are incorporated herein). In brief, Somata processes DNA samples, packages retroviral vector DNA into virions, transduces mammalian cells, and executes fluorescence-based bioassays, all in a 96-well format. In the case where the bioassay is a cell-lethal assay, the readout uses a membrane-impermeant dye (Sytox) to stain and detect cells that have a compromised membrane (i.e. dead/dying cells). Subsequently, a CCD camera image of each microplate well, in association with software designed to count stained cells, produced reliable measurements of dead/dying cells. Following each dead cell number count, co-treatment of the remaining cells with Sytox and a mild detergent (saponin), enabled the estimation of total cell number. By this method, the cytostatic properties of each agent could be assessed.

In addition to developing the above-described Somata assay, new vectors were constructed to test the presumptive cytotoxic/cytostatic agents. To maximize expression level and hence sensitivity of the assay, the HT29 cell line was modified with an expression cassette that encoded the HIV Tat gene product. To accomplish this, the pVT313 construct was digested with BamHI and ClaI to remove the fragment carrying the CMV promoter and adjacent eGFP coding sequence. The resulting pVT313 backbone was then gel purified and ligated to a 1.24kB BamHI/ClaI fragment (excised from pHi2-eGFP) carrying the CMV promoter operably linked to the Tat coding sequence. The resulting vector pVT1542, was then transformed into HT29 cells and selected to identify stable inserts. The nucleotide sequences encoding each presumptive cytotoxic/cytostatic agents were then cloned into an HIV2-based retroviral expression vector for testing in the Tat-modified HT29 cell line. Using techniques common to the field, the perturbagen insert was spliced into the EcoRI site of pVT1567. As a result of these manipulations each agent is attached to the C-terminus of a dead GFP (dGFP) scaffold which was, in turn, operably linked to an HIV2 promoter. When introduced into Tat-HT29 cell line, perturbagen expression is driven by the product of the Tat gene. As a result of these changes (in combination with the high MOI/high gene dosage promoted by Somata) transduced gene expression levels can reach as high as 30μM.

Using Somata in combination with the cell-lethal assay and Tat-engineered HT29 cells, 3,840 independent clones from the cycle 4 sublibrary and a cycle 7 sublibrary that had been depleted of Bid clones, were tested. The backgrounds in the assay were observed to be low and the reproducibility was observed to be high, thus allowing individual clones with dead cell numbers that exceeded the mean by two standard deviations (σ) and/or clones with total cell numbers less than 2σ, to be easily identified as candidate cell-lethal clones. Using the procedures described above, one hundred and nineteen clones were confirmed as producing *bona fide* cytotoxic/cytostatic effects. When DNA sequences were obtained for these clones and analyzed, the number condensed to a total of 11 unique sequences (see Figures 9-11; SEQ ID NOS: 3-13).

Sequence data showed that in addition to the two Bid clone types, 8 of 11 remaining sequences encoded products that were in-frame fusions of known or predicted native proteins (see Figures 10 and 11; SEQ ID NOS: 4-13). These known proteins included fragments of serum amyloid A (SEQ ID NO: 5), Arhgdig (SEQ ID NO: 6), MGC (SEQ ID NO: 8), and mRNA 24574 (SEQ ID NO: 4). Another sequence (1F1, contig of 3) exhibited >95% homology with a human sequence [NIH accession # BC002905 (SEQ ID NO: 12)]. The nucleotide sequence encoding the 1F1 perturbagen (SEQ ID NO: 13) stretches 1.415 kB,

and encodes an N-terminal truncated fragment of the BC002905 clone that is 396 amino acids in length. Interestingly, another clone encoded a long peptide (227 amino acids) derived from an out-of-frame translation of a β-tubulin cDNA [see 136C03 (SEQ ID NO: 7)]. Still another clone, 74A05 (SEQ ID NO: 9), produced a predicted fusion including an entire open reading frame of 154 amino acids derived from the 3' untranslated region of the RAB5C gene. Finally, one clone (138E10; SEQ ID NO: 10) encoded a predicted two-residue peptide fused to the C-terminus of GFP.

Because of the high frequency of BID and the concern of potential contamination by this strong cytotoxic agent, two methods were used to verify that individual clones were *bona fide* cell-lethals unrelated to BID. First, PCR reactions with Bid primers tested the possibility of low-level Bid clone contamination. Second, each unique clone was re-transformed into *E. coli*, and two independent colonies were picked, sequenced, and tested in Somata. Experiments with four replicates of each clone provided values for the kill index of the clones where the term "kill index" is defined as a normalized ratio of dead/dying cells to total cells in each well.

The kill index of the various cell-lethal clones varied from 20.3 to 1.3 (see Figure 12). The perturbagen having the most potent kill index was found to be the 120 amino acid fragment of MGC:2198 (SEQ ID NO: 8). In contrast, another perturbagen, 72D06, had a kill index of roughly 1, suggesting that this clone likely caused cell cycle arrest, rather than cell death.

Cell-lethal clones function at the protein level

To begin to address mechanistic questions about the cell-lethal clones, the four most penetrant clones (excluding Bid) were examined for activity at the protein level. Expression constructs in which a stop codon was inserted between the GFP scaffold sequence and the cDNA, were tested in Somata side-by-side with the parental constructs. In every case (including the two-residue peptide, 138E10; SEQ ID NO: 10), the altered constructs were far less active than their parents (see Figure 13). Northern blots confirmed that RNA expression in the constructs containing the added stop codon were roughly equivalent to RNA levels in cells transduced with parenteral constructs (data not shown). These experiments suggested that the activity of five cell-lethal clones resides in the protein, rather than the RNA, expressed products.

Selectively of Cytotoxic Clones

As a first step toward exploring phenotypic selectivity among the cell lethal clones, the cytotoxic effects of the perturbagens were tested in a second human colon

adenocarcinoma line. Using techniques described previously, SW620 cells (ATCC # CCL227) were engineered to carry a Tat expression construct. Subsequently, the cells were infected with the P_{HIV2}-perturbagen vectors, and selected to identify stable integrants. Cells carrying the two constructs were then tested in the context of Somata assays to assess and compare the cytotoxicity of the clones in the SW620 and HT29 backgrounds. Kill indices for the set of clones revealed that two clones, 138E10 (SEQ ID NO: 10) and 139G04 (SEQ ID NO: 11), were remarkably selective in their cytotoxic effects. 139G04 (SEQ ID NO: 11) had a kill index 11.5-fold higher in HT29 cells (see Figure 12). Similarly, 138E10 (SEQ ID NO: 10) exhibited a kill index that was 7.3-fold higher in HT29 cells than in SW620 cells. Other clones had equivalent indices in both cell types or mild (e.g., two-fold) preferences in one cell type or the other.

What is Claimed is:

1. A method of inducing cytotoxicity in a cell comprising the step of contacting said cell with an amount of a peptide fragment effective to induce cytotoxicity, said peptide fragment selected from the group consisting of: SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, and SEQ ID NO: 13.

2. A method of inducing cytotoxicity in a cell comprising the step of introducing into said cell a polynucleotide encoding a peptide fragment in an amount effective to induce cytotoxicity, said peptide fragment selected from the group consisting of: SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, and SEQ ID NO: 13.

3. The method of claims 1 or 2, wherein said cell is a mammalian cell.

4. The method of claim 3, wherein said cell is a cancer cell.

5. The method of claim 4, wherein said cancer cell is derived from a solid tumor.

6. The method of claim 4, wherein said cancer cell is metastatic.

7. The method of claim 4, wherein said cancer cell is derived from tissue selected from the group consisting of: breast, colon, lung, brain, melanoma, and prostate.

8. The method of claim 3, wherein said cell is a primary cell.

9. The method of claim 8, wherein said primary cell is selected from the group consisting of: epithelial cells, endothelial cells, stem cells, mesenchymal cells, fibroblasts, neuronal cells, and hematopoietic cells.

10. A peptide fragment consisting of the amino acid sequence of SEQ ID NO: 4.

11. A peptide fragment consisting of the amino acid sequence of SEQ ID NO: 5.

12. A peptide fragment consisting of the amino acid sequence of SEQ ID NO: 6.

13. A peptide fragment consisting of the amino acid sequence of SEQ ID NO: 7.

5

14. A peptide fragment consisting of the amino acid sequence of SEQ ID NO: 8.

15. A peptide fragment consisting of the amino acid sequence of SEQ ID NO: 9.

10 16. A peptide fragment consisting of the amino acid sequence of SEQ ID NO: 10.

17. A peptide fragment consisting of the amino acid sequence of SEQ ID NO: 11.

18. A peptide fragment consisting of the amino acid sequence of SEQ ID NO: 12.

15

19. A peptide fragment consisting of the amino acid sequence of SEQ ID NO: 13.

20. A polynucleotide encoding the peptide fragment of any one of claims 10 through 19.

Figure 1

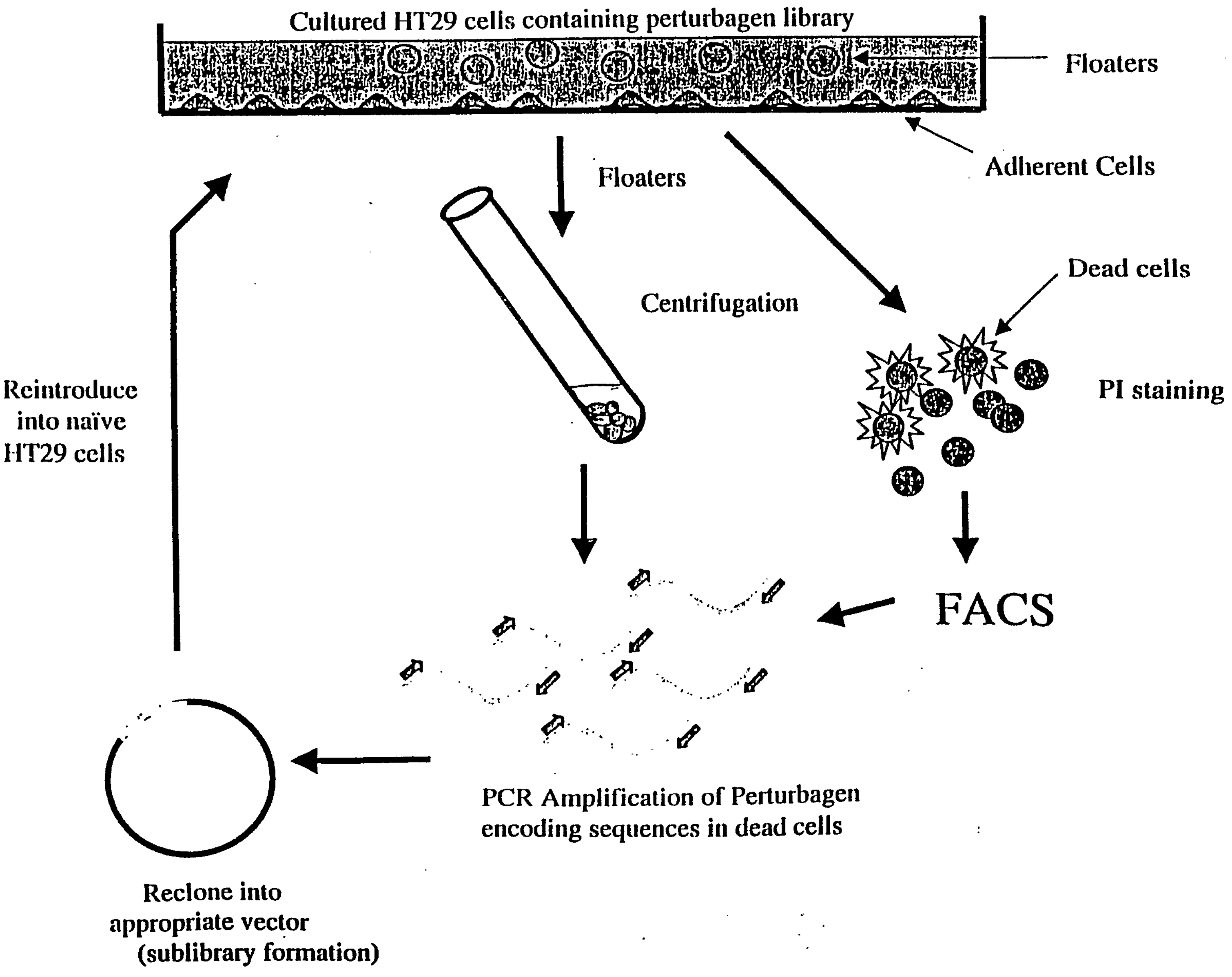

Figure 2

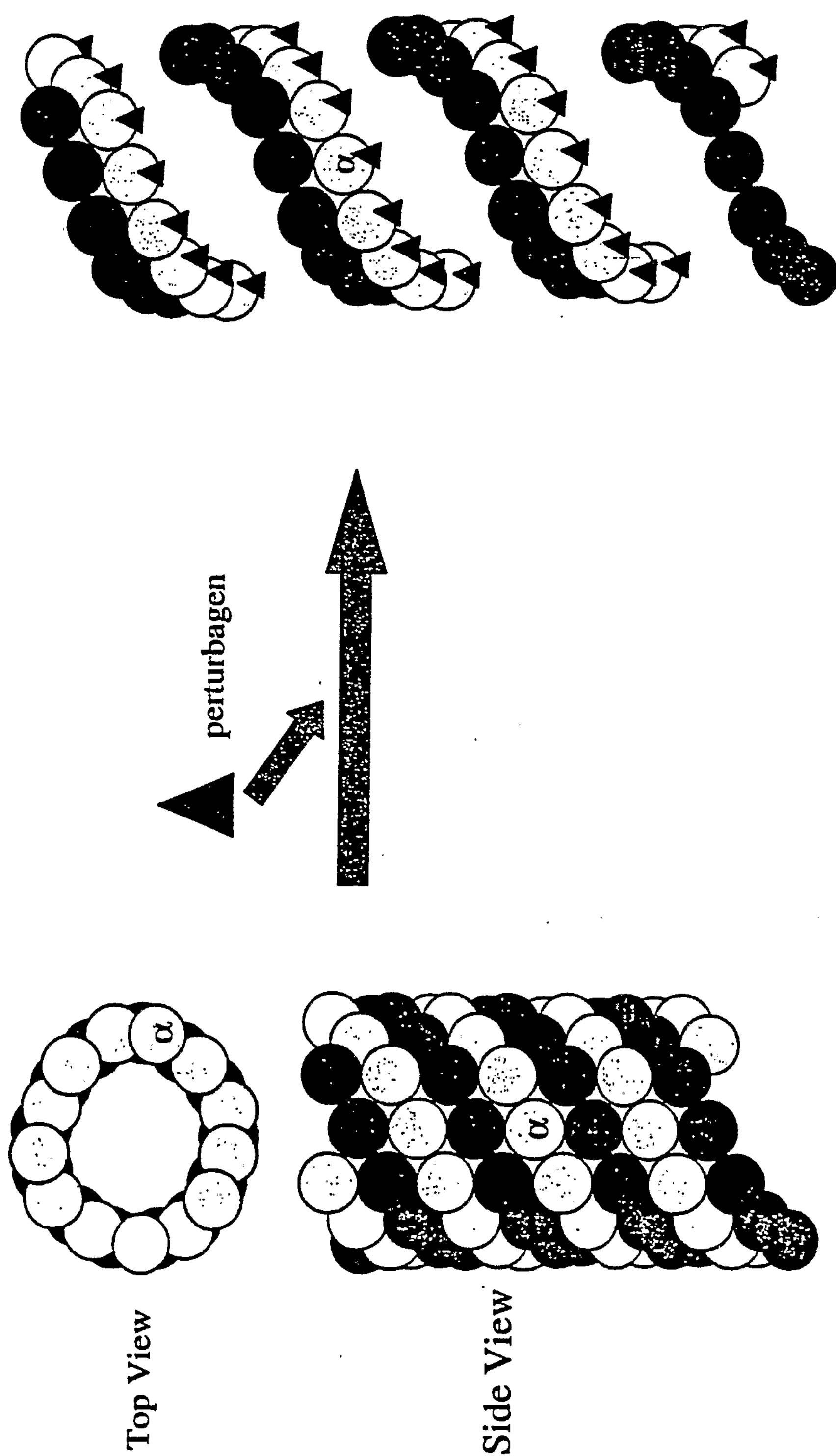

Figure 3

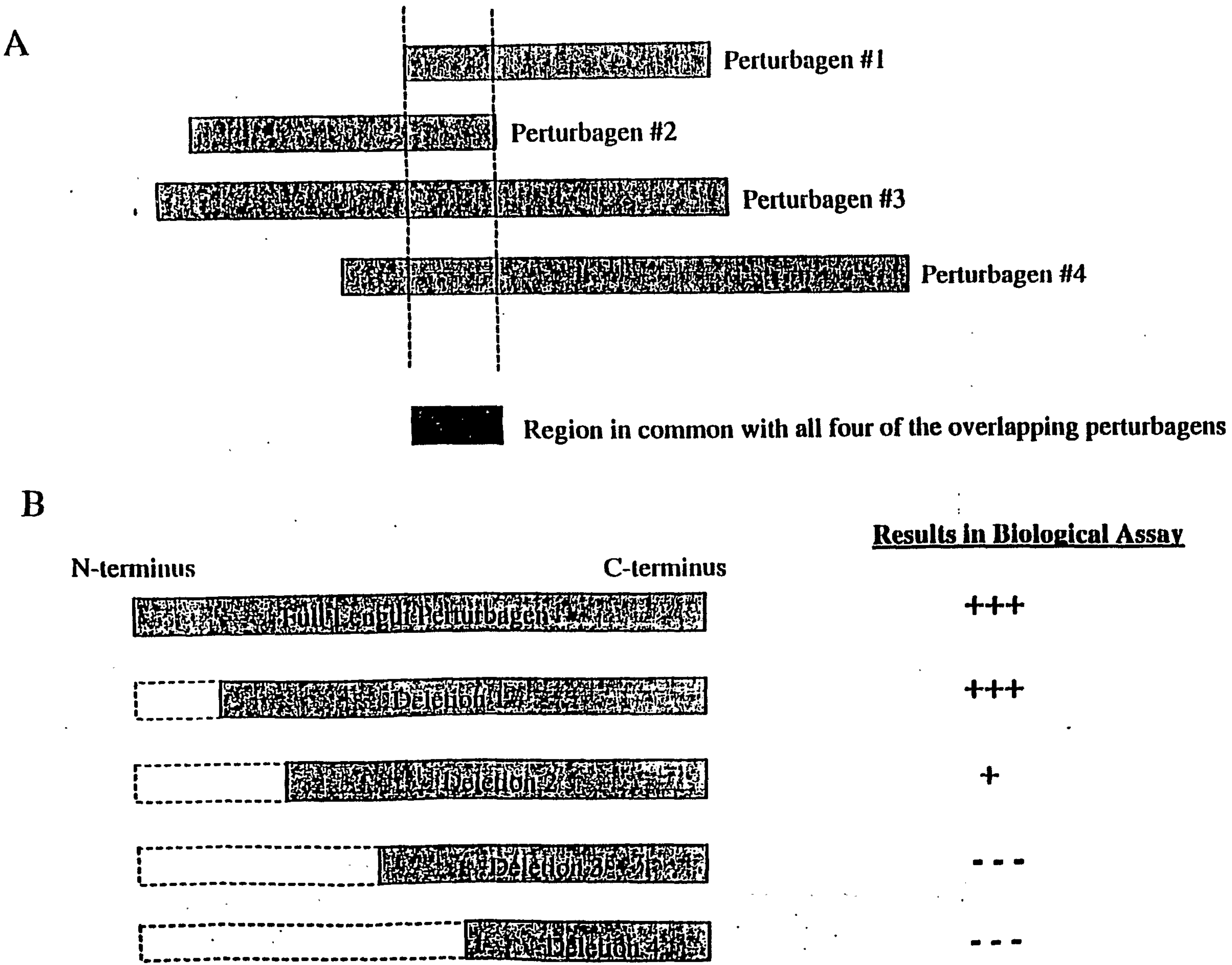

Figure 4

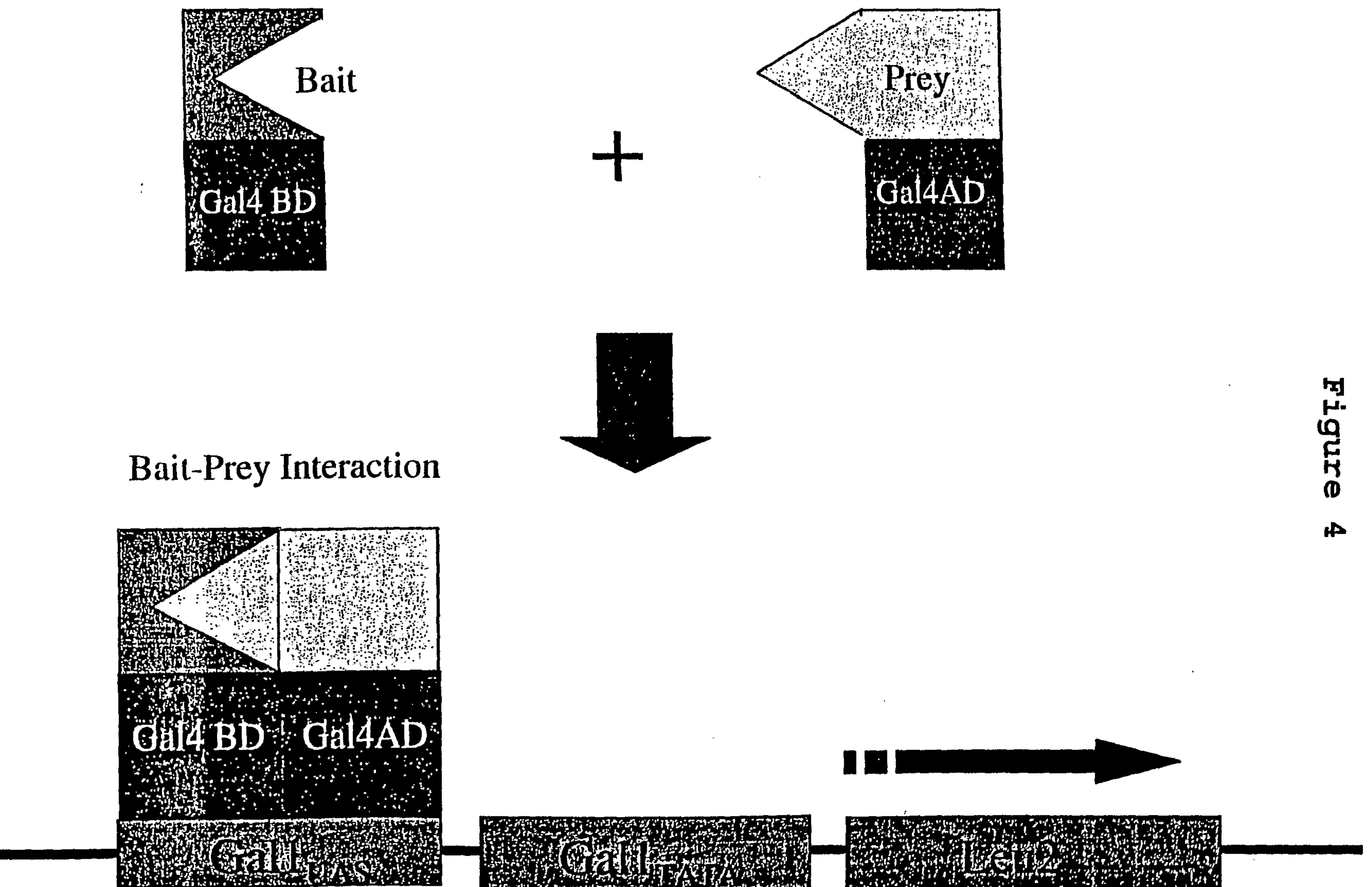

Figure 5

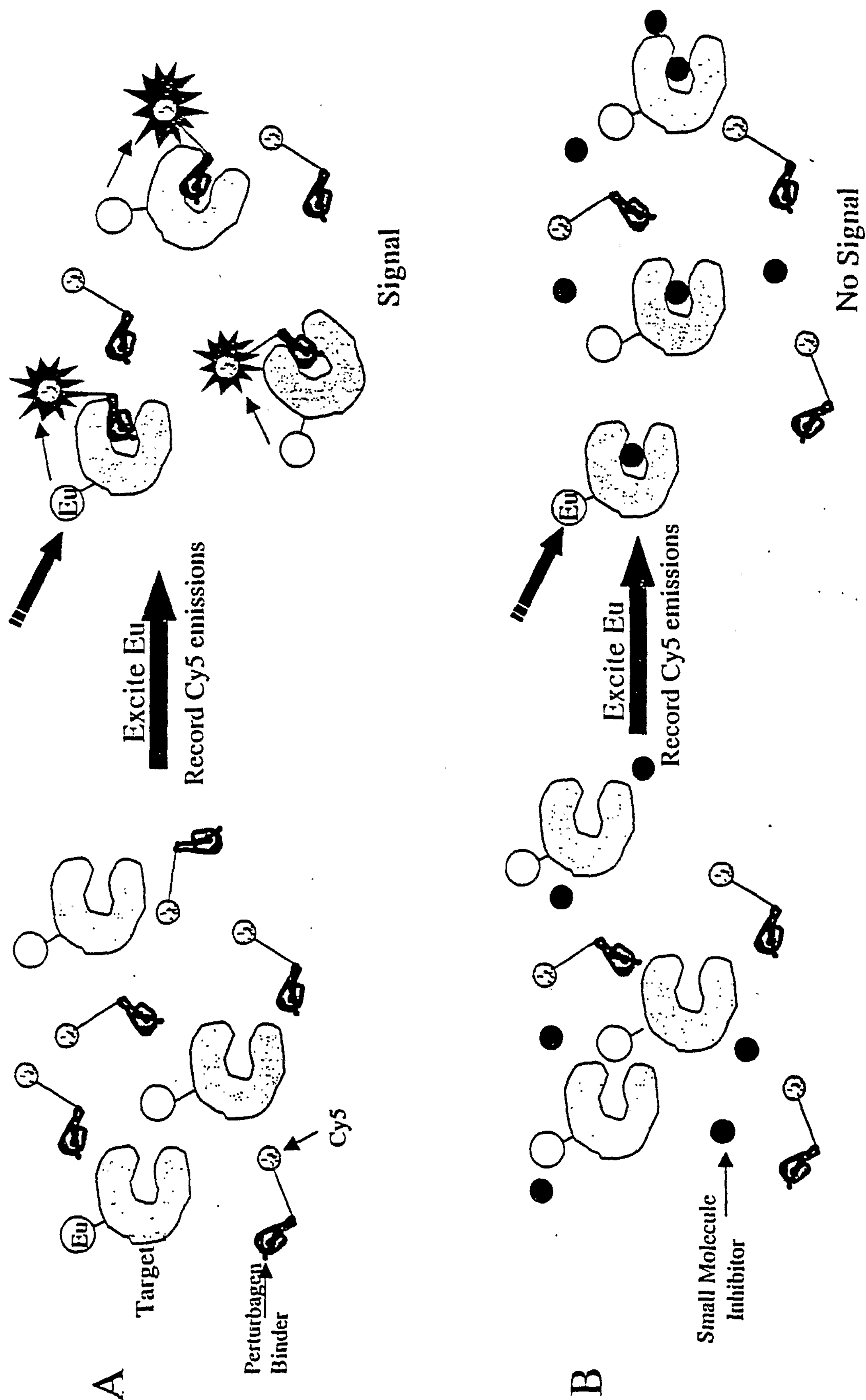

Figure 6

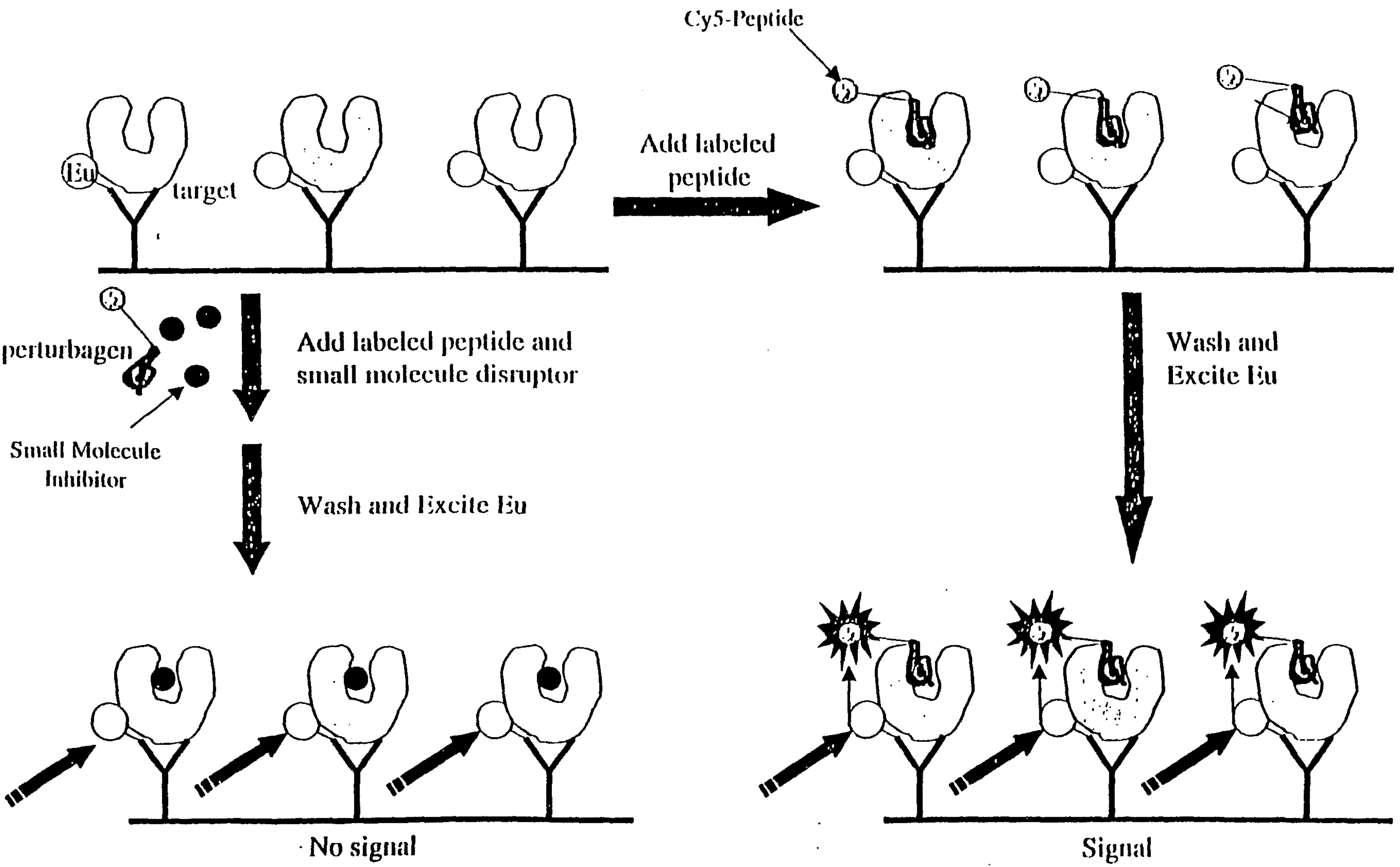

Figure 7

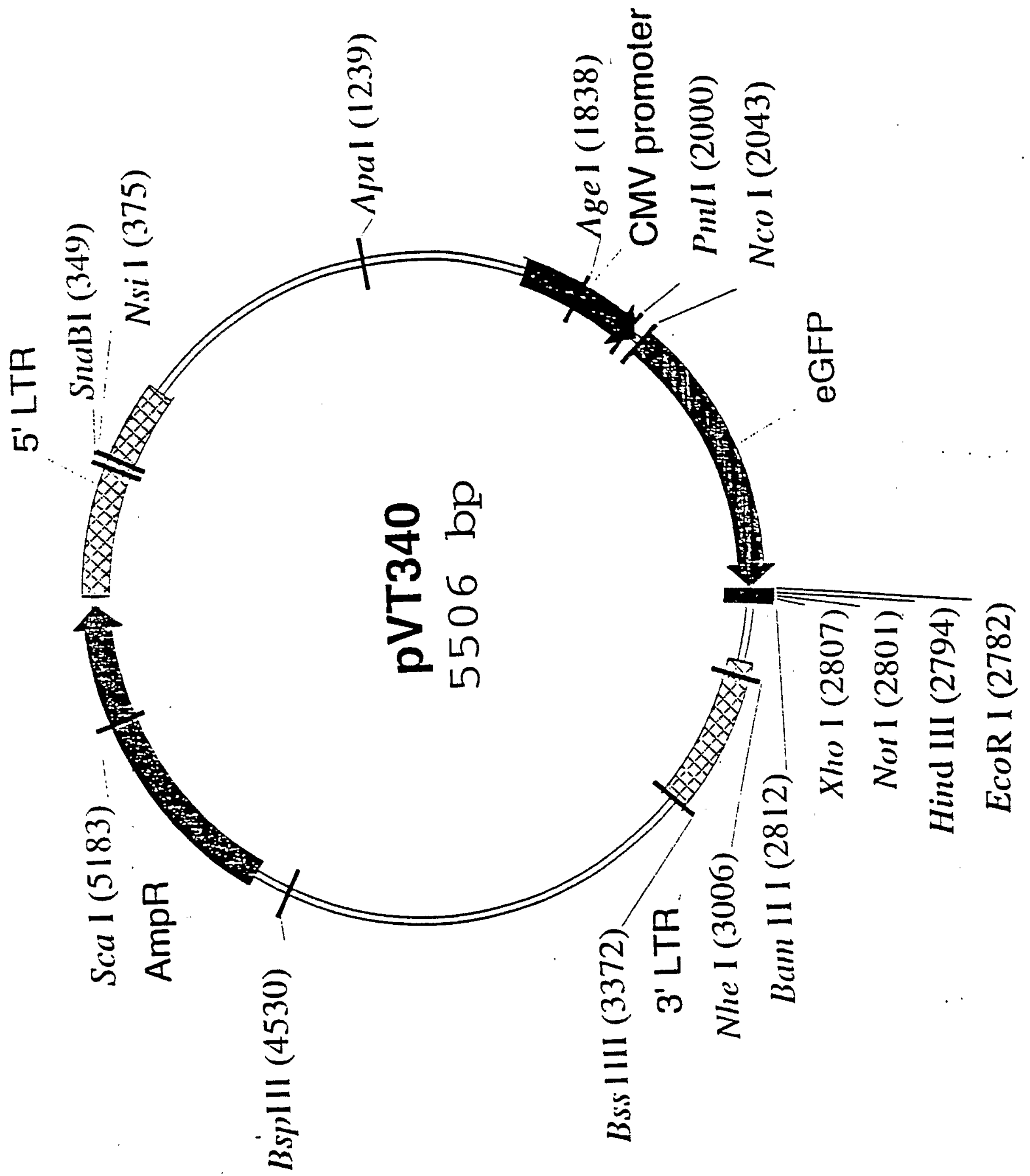

Figure 8

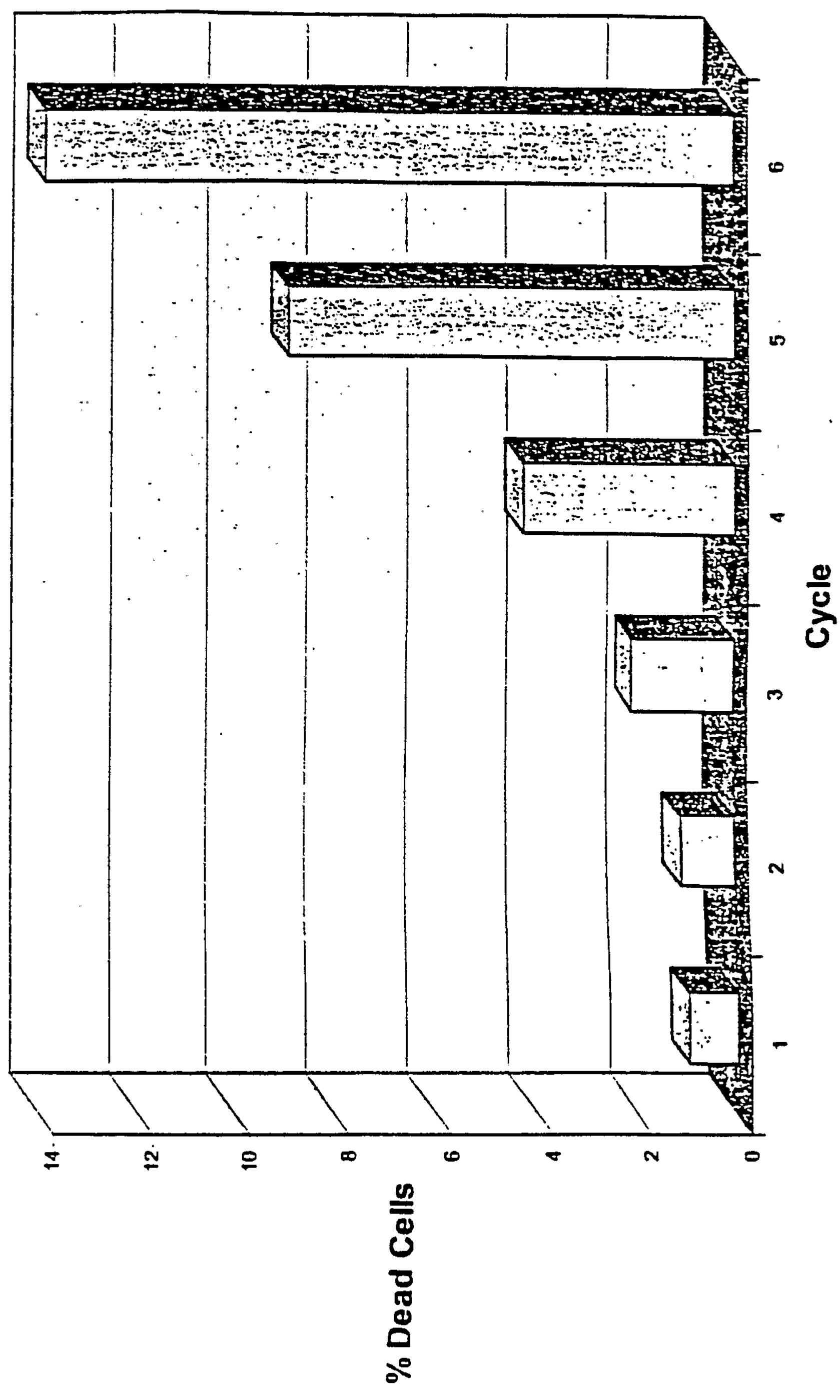

Figure 9

Sequence of BID (SEQ ID NO 3)

MDCEVNNGSSLRDECITNLLVFGFLQ
SCSDNSFRRELDALGHELPVLAPQWE
GYDELQTDGNRSSHSRLGRIEADSES
QEDIIRNIARHLAQVGDSMDRSIPPGL
VNGLALQLRNTSRSEEDRNRDLATAL
EQLLQAYPRDMEKEKTMLVLALLLAK
KVASHTPSLLRDVFHTTVNFINQNLRTY
VRSLARNGMD

Figure 10

Peptide Sequence of Perturbagens

72EO4-24574 mRNA (SEQ ID NO 4)
ASEGLGGRGHAPMQKAQKLPLRQALRLRSLRLPPYRLTSQEGQACIPQ
RRNCQTLSPVLGWLALGRASRVQPLLMALGLQFFHMCAPPAWPTLSLA
VGPEASSLPLGVSGIGMSAWLPSPPHLLLLSAAAGSGASHLRALGSSALE
GLQDPSQ

72HO4-Serum Amyloid (SEQ ID NO 5)
ALSFLGEAFDGARDMWRAYSDMREANYIGSDKYFHARGNYDAAKRGPG
GAWAAEVITDARENIQRFFGHGAEDSLADQAANEWGRSGKDPNHFRPAG
LPEKY

76FO3-ARHGDIG (SEQ ID NO 6)
EYRAPGRKSLLEIRQLDPDDRSLAKYKRVLLGPLPPAVDPSLPNVQVTRLTL
LSEQAPGPVVMDLTGDLAVLKDQVFVLKEGVDYRVKISFKVHREIVSGLKCL
HHTYRRGLRVDKTVYMVGSYGPSAQEYEFVTPVEEAPRGALVRGPYLVVS
LFTDDDRTHHLSWEWGLCICQDWKD

136C03-β-Tubulin (OF) (SEQ ID NO 7)
GAHGARAHPADVRLQEHDGRLRPAPRPLPDGGCHLPGPHVHEGGGRADA
QRAEQEQQLLRGVDPQQREDGRVRHPAPRPEDVGHLHRQQHGHPGAVQA
HLRAVHGHVPAQGLPALVHGRGHGRDGVHRGREQHERPGVRVPAVPGRH
GRRTRGVRGGGGRGRGVDAPARRVRESGGGKRGGGGLPGTITWQWKER
SMVYFRCALGLWCSSLLPVTFFSFFVILMTSM

144C05-MGC:2198 (SEQ ID NO 8)
EVGVPIFSCCCGPDMATPGPVIPEVPFEPSKPPVIEGLSPTVYRNPESFKEKF
VRKTRENPVVPIGCLATAAALTYGLYSFHRGNSQRSQLMMRTRIAAQGFTVA
AILLGLAVTAMKSRP

74A05-RAB5C (SEQ ID NO 9)
SSHMLLPPNPCIWGQWGQCSYPFFPSCGTADPARASTSSLCSEWSLVWAV
GQTLGRGLGRERQLFLQLALIRLQPPPRSHLPAGKPQHYHSIIVTATNPLPTTP
PPSVTPTSGSEPCSDQIMMMSIWGWVGKGGSGRGRNQLFLYFVLYVFFNM

138E10-Unknown (SEQ ID NO 10)
LQ

139G04-ACO22154, partial sequence (SEQ ID NO 11)
PQTQPAGPGLXTLGLLLLSLVPASPRPSGTLSCLILPAFPFNTAWSCVFQGLSR
HLLGSMQFTGLCQP

Figure 11

Protein Alignment of 1F1 Perturbagen and Wild Type BC002905 Sequence

```
MGERKGVNKYYPPDFNPEKHGSLNRYHNSHPLRERARKLSQGILIIRFEMPYNIWCDG

CKNHIGMGVRYNAEKKKVGNYYTTPIYRFRMKCHLCVNYIEMQTDPANCDYVIVSG

AQRKEERWDMADNEQVLTTEHEKKQKLETDAMFRLEHGEADRSTLKKALPTLSHIQ
                   EHEKKQKLETDAMFRLEHGEADRSTLKKALPTLSHIQ

EAQSAWKDDFALNSMLRRRFREKKKAIQEEEERDQALQAKASLTIPLVPETEDDRKL
EARSAWKDDFALNSMLRRRFREKKKAIQEEEERDQALQAKASLTIPLVPETEDDRKL

AALLKFHTLDSYEDKQKLKRTEIISRSWFPSAPGSASSSKVSGVLKKLAQSRRTALATS
AALLKFHTLDSYEDKQKLKRTEIISRSWFPSAPGSASSSKVSDVLKKLAQSRRTAPATS

PITVGDLGIVRRRSRDVPESPQHAADTPKSGEPRVPEEAAQDRPMSPGDCPPETTETPK
PITVGDLGIVRRRSRDVPESPQHAADTPKSGEPRVPEEAAQDRPMSPGDCPPEATETPK

CSSPRGQEGSRQDKPLSPAGSSQEAADTPDTRHPCSLGSSLVADYSDSESE (SEQ ID NO 12)
CSSPRGQEGSRQDKPLSPAGSSQEAADTPDTRHPCSLGSSLVADYSDSESE (SEQ ID NO 13)
```

Figure 12

Kill Index of Perturbagens in HT29 and SW620 Cells

	Kill Index (HT29)	Kill Index (SW620)
pVT1567	1.0	1.0
BidA	19.8	--
BidB	21.9	14.6
F1	3.0	3.8
Serum Amyloid	20.1	10.2
24574	1.3	1.3
ARHGDIG	2.0	2.0
138E1	17.5	2.4
B-tubulin	2.2	4.7
ACO22154	18.4	1.6
MGC:2198	20.3	6.6
RAB5C	1.4	2.6

Figure 13

Kill Index of Perturbagens In-Frame and Out-of-Frame

	Kill Index (IF)	Kill Index (OF)
pVT1567	1.0	0.9
Serum Amyloid	20.1	0.9
138E1	17.5	1.3
ACO22154	18.4	1.8
MGC:2198	20.3	0.7

SEQUENCE LISTING

<110> Deltagen Proteomics Inc.

<120> Negative Selections Assays, And Compositions Thereof

<130> 29345/38668A

<160> 13

<170> PatentIn version 3.2

<210> 1
<211> 18
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic primer

<400> 1
gccgccggga tcactctc 18

<210> 2
<211> 27
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic primer

<400> 2
gctagcttgc caaacctaca ggtgggg 27

<210> 3
<211> 195
<212> PRT
<213> Homo sapiens

<400> 3

Met Asp Cys Glu Val Asn Asn Gly Ser Ser Leu Arg Asp Glu Cys Ile
1 5 10 15

Thr Asn Leu Leu Val Phe Gly Phe Leu Gln Ser Cys Ser Asp Asn Ser
20 25 30

Phe Arg Arg Glu Leu Asp Ala Leu Gly His Glu Leu Pro Val Leu Ala
35 40 45

Pro Gln Trp Glu Gly Tyr Asp Glu Leu Gln Thr Asp Gly Asn Arg Ser
50 55 60

Ser His Ser Arg Leu Gly Arg Ile Glu Ala Asp Ser Glu Ser Gln Glu
65 70 75 80

```
Asp Ile Ile Arg Asn Ile Ala Arg His Leu Ala Gln Val Gly Asp Ser
                85                  90                  95

Met Asp Arg Ser Ile Pro Pro Gly Leu Val Asn Gly Leu Ala Leu Gln
            100                 105                 110

Leu Arg Asn Thr Ser Arg Ser Glu Glu Asp Arg Asn Arg Asp Leu Ala
        115                 120                 125

Thr Ala Leu Glu Gln Leu Leu Gln Ala Tyr Pro Arg Asp Met Glu Lys
    130                 135                 140

Glu Lys Thr Met Leu Val Leu Ala Leu Leu Leu Ala Lys Lys Val Ala
145                 150                 155                 160

Ser His Thr Pro Ser Leu Leu Arg Asp Val Phe His Thr Thr Val Asn
                165                 170                 175

Phe Ile Asn Gln Asn Leu Arg Thr Tyr Val Arg Ser Leu Ala Arg Asn
            180                 185                 190

Gly Met Asp
        195


<210>  4
<211>  153
<212>  PRT
<213>  Homo sapiens

<400>  4

Ala Ser Glu Gly Leu Gly Gly Arg Gly His Ala Pro Met Gln Lys Ala
1               5                   10                  15

Gln Lys Leu Pro Leu Arg Gln Ala Leu Arg Leu Arg Ser Leu Arg Leu
            20                  25                  30

Pro Pro Tyr Arg Leu Thr Ser Gln Glu Gly Gln Ala Cys Ile Pro Gln
        35                  40                  45

Arg Arg Asn Cys Gln Thr Leu Ser Pro Val Leu Gly Trp Leu Ala Leu
    50                  55                  60

Gly Arg Ala Ser Arg Val Gln Pro Leu Leu Met Ala Leu Gly Leu Gln
65                  70                  75                  80

Phe Phe His Met Cys Ala Pro Pro Ala Trp Pro Thr Leu Ser Leu Ala
                85                  90                  95
```

```
Val Gly Pro Glu Ala Ser Ser Leu Pro Leu Phe Val Ser Gly Ile Gly
            100                 105                 110

Met Ser Ala Trp Leu Pro Ser Pro Pro His Leu Leu Leu Leu Ser Ala
        115                 120                 125

Ala Ala Gly Ser Gly Ala Ser His Leu Arg Ala Leu Gly Ser Ser Ala
    130                 135                 140

Leu Glu Gly Leu Gln Asp Pro Ser Gln
145                 150

<210>  5
<211>  102
<212>  PRT
<213>  Homo sapiens

<400>  5

Ala Leu Ser Phe Leu Gly Glu Ala Phe Asp Gly Ala Arg Asp Met Trp
1               5                   10                  15

Arg Ala Tyr Ser Asp Met Arg Glu Ala Asn Tyr Ile Gly Ser Asp Lys
            20                  25                  30

Tyr Phe His Ala Arg Gly Asn Tyr Asp Ala Ala Lys Arg Gly Pro Gly
        35                  40                  45

Gly Ala Trp Ala Ala Glu Val Ile Thr Asp Ala Arg Glu Asn Ile Gln
    50                  55                  60

Arg Phe Phe Gly His Gly Ala Glu Asp Ser Leu Ala Asp Gln Ala Ala
65                  70                  75                  80

Asn Glu Trp Gly Arg Ser Gly Lys Asp Pro Asn His Phe Arg Pro Ala
                85                  90                  95

Gly Leu Pro Glu Lys Tyr
            100


<210>  6
<211>  179
<212>  PRT
<213>  Homo sapiens

<400>  6

Glu Tyr Arg Ala Pro Gly Arg Lys Ser Leu Leu Glu Ile Arg Gln Leu
1               5                   10                  15
```

```
Asp Pro Asp Asp Arg Ser Leu Ala Lys Tyr Lys Arg Val Leu Leu Gly
            20                  25                  30

Pro Leu Pro Pro Ala Val Asp Pro Ser Leu Pro Asn Val Gln Val Thr
        35                  40                  45

Arg Leu Thr Leu Leu Ser Glu Gln Ala Pro Gly Pro Val Val Met Asp
    50                  55                  60

Leu Thr Gly Asp Leu Ala Val Leu Lys Asp Gln Val Phe Val Leu Lys
65                  70                  75                  80

Glu Gly Val Asp Tyr Arg Val Lys Ile Ser Phe Lys Val His Arg Glu
                85                  90                  95

Ile Val Ser Gly Leu Lys Cys Leu His His Thr Tyr Arg Arg Gly Leu
            100                 105                 110

Arg Val Asp Lys Thr Val Tyr Met Val Gly Ser Tyr Gly Pro Ser Ala
        115                 120                 125

Gln Glu Tyr Glu Phe Val Thr Pro Val Glu Glu Ala Pro Arg Gly Ala
    130                 135                 140

Leu Val Arg Gly Pro Tyr Leu Val Val Ser Leu Phe Thr Asp Asp Asp
145                 150                 155                 160

Arg Thr His His Leu Ser Trp Glu Trp Gly Leu Cys Ile Cys Gln Asp
                165                 170                 175

Trp Lys Asp


<210>  7
<211>  227
<212>  PRT
<213>  Homo sapiens

<400>  7

Gly Ala His Gly Ala Arg Ala His Pro Ala Asp Val Arg Leu Gln Glu
1               5                   10                  15

His Asp Gly Arg Leu Arg Pro Ala Pro Arg Pro Leu Pro Asp Gly Gly
            20                  25                  30

Cys His Leu Pro Gly Pro His Val His Glu Gly Gly Gly Arg Ala Asp
        35                  40                  45
```

Ala Gln Arg Ala Glu Gln Glu Gln Gln Leu Leu Arg Gly Val Asp Pro
50 55 60

Gln Gln Arg Glu Asp Gly Arg Val Arg His Pro Ala Pro Arg Pro Glu
65 70 75 80

Asp Val Gly His Leu His Arg Gln Gln His Gly His Pro Gly Ala Val
85 90 95

Gln Ala His Leu Arg Ala Val His Gly His Val Pro Ala Gln Gly Leu
100 105 110

Pro Ala Leu Val His Gly Arg Gly His Gly Arg Asp Gly Val His Arg
115 120 125

Gly Arg Glu Gln His Glu Arg Pro Gly Val Arg Val Pro Ala Val Pro
130 135 140

Gly Arg His Gly Arg Arg Thr Arg Gly Val Arg Gly Gly Gly Gly Arg
145 150 155 160

Gly Arg Gly Val Asp Ala Pro Ala Arg Arg Val Arg Glu Ser Gly Gly
165 170 175

Gly Lys Arg Gly Gly Gly Gly Leu Pro Gly Thr Ile Thr Trp Gln Trp
180 185 190

Lys Glu Arg Ser Met Val Tyr Phe Arg Cys Ala Leu Gly Leu Trp Cys
195 200 205

Ser Ser Leu Leu Pro Val Thr Phe Phe Ser Phe Phe Val Ile Leu Met
210 215 220

Thr Ser Met
225

<210> 8
<211> 120
<212> PRT
<213> Homo sapiens

<400> 8

Glu Val Gly Val Pro Ile Phe Ser Cys Cys Cys Gly Pro Asp Met Ala
1 5 10 15

Thr Pro Gly Pro Val Ile Pro Glu Val Pro Phe Glu Pro Ser Lys Pro
20 25 30

Pro Val Ile Glu Gly Leu Ser Pro Thr Val Tyr Arg Asn Pro Glu Ser
35 40 45

Phe Lys Glu Lys Phe Val Arg Lys Thr Arg Glu Asn Pro Val Val Pro
50 55 60

Ile Gly Cys Leu Ala Thr Ala Ala Ala Leu Thr Tyr Gly Leu Tyr Ser
65 70 75 80

Phe His Arg Gly Asn Ser Gln Arg Ser Gln Leu Met Met Arg Thr Arg
85 90 95

Ile Ala Ala Gln Gly Phe Thr Val Ala Ala Ile Leu Leu Gly Leu Ala
100 105 110

Val Thr Ala Met Lys Ser Arg Pro
115 120

<210> 9
<211> 154
<212> PRT
<213> Homo sapiens

<400> 9

Ser Ser His Met Leu Leu Pro Pro Asn Pro Cys Ile Trp Gly Gln Trp
1 5 10 15

Gly Gln Cys Ser Tyr Pro Phe Phe Pro Ser Cys Gly Thr Ala Asp Pro
20 25 30

Ala Arg Ala Ser Thr Ser Ser Leu Cys Ser Glu Trp Ser Leu Val Trp
35 40 45

Ala Val Gly Gln Thr Leu Gly Arg Gly Leu Gly Arg Glu Arg Gln Leu
50 55 60

Phe Leu Gln Leu Ala Leu Ile Arg Leu Gln Pro Pro Pro Arg Ser His
65 70 75 80

Leu Pro Ala Gly Lys Pro Gln His Tyr His Ser Ile Ile Val Thr Ala
85 90 95

Thr Asn Pro Leu Pro Thr Thr Pro Pro Pro Ser Val Thr Pro Thr Ser
100 105 110

Gly Ser Glu Pro Cys Ser Asp Gln Ile Met Met Met Ser Ile Trp Gly
115 120 125

Trp Val Gly Lys Gly Gly Ser Gly Arg Gly Arg Asn Gln Leu Phe Leu
    130   135   140

Tyr Phe Val Leu Tyr Val Phe Phe Asn Met
145   150

<210> 10
<211> 2
<212> PRT
<213> Homo sapiens

<400> 10

Leu Gln
1

<210> 11
<211> 68
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<222> (11)..(11)
<223> Xaa = any or unknown amino acid

<400> 11

Pro Gln Thr Gln Pro Ala Gly Pro Gly Leu Xaa Thr Leu Gly Leu Leu
1   5   10   15

Leu Leu Ser Leu Val Pro Ala Ser Pro Arg Pro Ser Gly Thr Leu Ser
            20   25   30

Cys Leu Ile Leu Pro Ala Phe Pro Phe Asn Thr Ala Trp Ser Cys Val
        35   40   45

Phe Gln Gly Leu Ser Arg His Leu Leu Gly Ser Met Gln Phe Thr Gly
    50   55   60

Leu Cys Gln Pro
65

<210> 12
<211> 396
<212> PRT
<213> Homo sapiens

<400> 12

Met Gly Glu Arg Lys Gly Val Asn Lys Tyr Tyr Pro Pro Asp Phe Asn
1   5   10   15

Pro Glu Lys His Gly Ser Leu Asn Arg Tyr His Asn Ser His Pro Leu
            20   25   30

```
Arg Glu Arg Ala Arg Lys Leu Ser Gln Gly Ile Leu Ile Ile Arg Phe
        35                  40                  45

Glu Met Pro Tyr Asn Ile Trp Cys Asp Gly Cys Lys Asn His Ile Gly
    50                  55                  60

Met Gly Val Arg Tyr Asn Ala Glu Lys Lys Lys Val Gly Asn Tyr Tyr
65                  70                  75                  80

Thr Thr Pro Ile Tyr Arg Phe Arg Met Lys Cys His Leu Cys Val Asn
                85                  90                  95

Tyr Ile Glu Met Gln Thr Asp Pro Ala Asn Cys Asp Tyr Val Ile Val
            100                 105                 110

Ser Gly Ala Gln Arg Lys Glu Glu Arg Trp Asp Met Ala Asp Asn Glu
        115                 120                 125

Gln Val Leu Thr Thr Glu His Glu Lys Lys Gln Lys Leu Glu Thr Asp
    130                 135                 140

Ala Met Phe Arg Leu Glu His Gly Glu Ala Asp Arg Ser Thr Leu Lys
145                 150                 155                 160

Lys Ala Leu Pro Thr Leu Ser His Ile Gln Glu Ala Gln Ser Ala Trp
                165                 170                 175

Lys Asp Asp Phe Ala Leu Asn Ser Met Leu Arg Arg Arg Phe Arg Glu
            180                 185                 190

Lys Lys Lys Ala Ile Gln Glu Glu Glu Glu Arg Asp Gln Ala Leu Gln
        195                 200                 205

Ala Lys Ala Ser Leu Thr Ile Pro Leu Val Pro Glu Thr Glu Asp Asp
    210                 215                 220

Arg Lys Leu Ala Ala Leu Leu Lys Phe His Thr Leu Asp Ser Tyr Glu
225                 230                 235                 240

Asp Lys Gln Lys Leu Lys Arg Thr Glu Ile Ile Ser Arg Ser Trp Phe
                245                 250                 255

Pro Ser Ala Pro Gly Ser Ala Ser Ser Ser Lys Val Ser Gly Val Leu
            260                 265                 270
```

Lys Lys Leu Ala Gln Ser Arg Arg Thr Ala Leu Ala Thr Ser Pro Ile
275 280 285

Thr Val Gly Asp Leu Gly Ile Val Arg Arg Arg Ser Arg Asp Val Pro
290 295 300

Glu Ser Pro Gln His Ala Ala Asp Thr Pro Lys Ser Gly Glu Pro Arg
305 310 315 320

Val Pro Glu Glu Ala Ala Gln Asp Arg Pro Met Ser Pro Gly Asp Cys
325 330 335

Pro Pro Glu Thr Thr Glu Thr Pro Lys Cys Ser Ser Pro Arg Gly Gln
340 345 350

Glu Gly Ser Arg Gln Asp Lys Pro Leu Ser Pro Ala Gly Ser Ser Gln
355 360 365

Glu Ala Ala Asp Thr Pro Asp Thr Arg His Pro Cys Ser Leu Gly Ser
370 375 380

Ser Leu Val Ala Asp Tyr Ser Asp Ser Glu Ser Glu
385 390 395

<210> 13
<211> 263
<212> PRT
<213> Homo sapiens

<400> 13

Glu His Glu Lys Lys Gln Lys Leu Glu Thr Asp Ala Met Phe Arg Leu
1 5 10 15

Glu His Gly Glu Ala Asp Arg Ser Thr Leu Lys Lys Ala Leu Pro Thr
20 25 30

Leu Ser His Ile Gln Glu Ala Gln Ser Ala Trp Lys Asp Asp Phe Ala
35 40 45

Leu Asn Ser Met Leu Arg Arg Arg Phe Arg Glu Lys Lys Lys Ala Ile
50 55 60

Gln Glu Glu Glu Glu Arg Asp Gln Ala Leu Gln Ala Lys Ala Ser Leu
65 70 75 80

Thr Ile Pro Leu Val Pro Glu Thr Glu Asp Asp Arg Lys Leu Ala Ala
85 90 95

```
Leu Leu Lys Phe His Thr Leu Asp Ser Tyr Glu Asp Lys Gln Lys Leu
            100                 105                 110

Lys Arg Thr Glu Ile Ile Ser Arg Ser Trp Phe Pro Ser Ala Pro Gly
        115                 120                 125

Ser Ala Ser Ser Ser Lys Val Ser Asp Val Leu Lys Lys Leu Ala Gln
    130                 135                 140

Ser Arg Arg Thr Ala Leu Ala Thr Ser Pro Ile Thr Val Gly Asp Leu
145                 150                 155                 160

Gly Ile Val Arg Arg Arg Ser Arg Asp Val Pro Glu Ser Pro Gln His
                165                 170                 175

Ala Ala Asp Thr Pro Lys Ser Gly Glu Pro Arg Val Pro Glu Glu Ala
            180                 185                 190

Ala Gln Asp Arg Pro Met Ser Pro Gly Asp Cys Pro Pro Glu Thr Thr
        195                 200                 205

Glu Thr Pro Lys Cys Ser Ser Pro Arg Gly Gln Glu Gly Ser Arg Gln
    210                 215                 220

Asp Lys Pro Leu Ser Pro Ala Gly Ser Ser Gln Glu Ala Ala Asp Thr
225                 230                 235                 240

Pro Asp Thr Arg His Pro Cys Ser Leu Gly Ser Ser Leu Val Ala Asp
                245                 250                 255

Tyr Ser Asp Ser Glu Ser Glu
            260
```