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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 May 2004 (13.05.2004)

PCT

(10) International Publication Number
WO 2004/039943 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US2003/015465

(22) International Filing Date: 16 May 2003 (16.05.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/381,533 17 May 2002 (17.05.2002) US
60/445,222 4 February 2003 (04.02.2003) US

(71) Applicant (*for all designated States except US*): **CHIRON CORPORATION** [US/US]; 4560 Horton Street - R440, Emeryville, CA 94608 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **SCOTT, Elizabeth, M.** [US/US]; 1267 Nash Street, Sonoma, CA 95476 (US). **LAMSON, George** [US/US]; 232 Sandringham Drive, Moraga, CA 94556 (US). **KASSAM, Altaf** [US/US]; 3810 Midvale Avenue, Oakland, CA 94602 (US). **ZHANG, Guozhong** [CN/US]; 41236 Norman Court, Fremont, CA 94539 (US). **SAKAMOTO, Doreen** [US/US]; 6655 Moore Drive, Oakland, CA 94611 (US). **GARCIA, Pablo, Dominguez** [CL/US]; 882 Chenery Street, San Francisco, CA 94131 (US).

(74) Agent: **BLACKBURN, Robert, P.**; Chiron Corporation, P.O. Box 8097, Emeryville, CA 94662-8097 (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, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, 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*
— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

*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.*

WO 2004/039943 A2

(54) **Title:** HUMAN GENES AND GENE EXPRESSION PRODUCTS ISOLATED FROM HUMAN PROSTATE

(57) **Abstract:** This invention relates to novel human polynucleotides and variants thereof, their encoded polypeptides and variants thereof, to genes corresponding to these polynucleotides and to proteins expressed by the genes. The invention also relates to diagnostics and therapeutics comprising such novel human polynucleotides, their corresponding genes or gene products, including probes, antisense nucleotides, and antibodies. The polynucleotides of the invention correspond to a polynucleotide comprising the sequence information of at least one of SEQ ID NOS:1-1485. The polypeptides of the invention correspond to a polypeptide comprising the amino acid sequence information of at least one of SEQ ID NOS:1486-1542.

# HUMAN GENES AND GENE EXPRESSION PRODUCTS ISOLATED FROM HUMAN PROSTATE

## Field of the Invention

The present invention relates to polynucleotides of human origin, particularly in human prostate, and the encoded gene products.

## Background of the Invention

Identification of novel polynucleotides, particularly those that encode an expressed gene product, is important in the advancement of drug discovery, diagnostic technologies, and the understanding of the progression and nature of complex diseases such as cancer. Identification of genes expressed in different cell types isolated from sources that differ in disease state or stage, developmental stage, exposure to various environmental factors, the tissue of origin, the species from which the tissue was isolated, and the like is key to identifying the genetic factors that are responsible for the phenotypes associated with these various differences.

This invention provides novel human polynucleotides, the polypeptides encoded by these polynucleotides, and the genes and proteins corresponding to these novel polynucleotides.

## Summary of the Invention

This invention relates to novel human polynucleotides and variants thereof, their encoded polypeptides and variants thereof, to genes corresponding to these polynucleotides and to proteins expressed by the genes. The invention also relates to diagnostics and therapeutics comprising such novel human polynucleotides, their corresponding genes or gene products, including probes, antisense nucleotides, and antibodies. The polynucleotides of the invention correspond to a polynucleotide comprising the sequence information of at least one of SEQ ID NOS:1-1485. The polypeptides of the invention correspond to a polypeptide comprising the amino acid sequence information of at least one of SEQ ID NOS:1486-1542.

Accordingly, in one aspect, the invention provides an isolated polynucleotide comprising a nucleotide sequence which hybridizes under stringent conditions to a sequence selected from the group consisting of SEQ ID NOS: 1-1485.

In another aspect, the invention provides an isolated polynucleotide comprising at least 15 contiguous nucleotides of a nucleotide sequence having at least 90% sequence identity to a sequence selected from the group consisting of SEQ ID NOS:1-1485, a degenerate variant of SEQ ID NOS:1-1485, an antisense of SEQ ID NOS:1-1485, and a complement of SEQ ID NOS:1-1485.

In another aspect, the invention provides an isolated polynucleotide comprising at least 15 contiguous nucleotides of a nucleotide sequence selected from the group consisting of SEQ ID NOS:1-1485, a degenerate variant of SEQ ID NOS:1-1485, an antisense of SEQ ID NOS:1-1485, and a complement of SEQ ID NOS:1-1485. In specific embodiments, the polynucleotide comprises at least 100 contiguous nucleotides of the nucleotide sequence. In other specific embodiments, the poynucleotide comprises at least 200 contiguous nucleotides of the nucleotide sequence.

In another aspect, the invention provides An isolated polynucleotide comprising a nucleotide sequence of at least 90% sequence identity to a sequence selected from the group consisting of SEQ ID NOS:1-1485, a degenerate variant of SEQ ID NOS:1-1485, an antisense of SEQ ID NOS:1-1485, and a complement of SEQ ID NOS:1-1485. In specific embodiments, the polynucleotide comprises a nucleotide sequence of at least 95% sequence identity to the selected nucleotide sequence. In other specific embodiments, the polynucleotide comprises a nucleotide sequence that is identical to the selected nucleotide sequence.

In another aspect, the invention provides a polynucleotide comprising a nucleotide sequence of an insert contained in a clone deposited as NRRL Accession No. B-30523, B-30524, B-30525, B-30526, B-30527, B-30528, B-30529, or B-30581.

In another aspect, the invention provides an isolated cDNA obtained by the process of amplification using a polynucleotide comprising at least 15 contiguous nucleotides of a nucleotide sequence selected from the group consisting of SEQ ID NOS:1-1485. In specific embodiments, the polynucleotide comprises at least 25 contiguous nucleotides of the selected nucleotide sequence. In other specific embodiments, the polynucleotide comprises at least 100 contiguous nucleotides of the selected nucleotide sequence. In some embodiments, the amplification is by polymerase chain reaction (PCR) amplification.

In another aspect, the invention provides an isolated recombinant host cell containing a polynucleotide of the invention.

In another aspect, the invention provides an isolated vector comprising a polynucleotide of the invention.

In another aspect, the invention provides a method for producing a polypeptide, the method comprising the steps of culturing a recombinant host cell containing a polynucleotide of the invention under conditions suitable for the expression of an encoded polypeptide and recovering the polypeptide from the host cell culture.

In another aspect, the invention provides an isolated polypeptide encoded by a poynucleotide of the invention.

In another aspect, the invention provides an isolated polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOS:1486-1542.

In another aspect, the invention provides an antibody that specifically binds a polypeptide of the invention.

In another aspect, the invention provides a method of detecting differentially expressed genes correlated with a cancerous state of a mammalian cell, the method comprising the step of detecting at least one differentially expressed gene product in a test sample derived from a cell suspected of being cancerous, where the gene product is encoded by a gene comprising an identifying sequence of at least one of SEQ ID NOS:1-1485. Detection of the differentially expressed gene product is correlated with a cancerous state of the cell from which the test sample was derived.

In another aspect, the invention provides a method of detecting differentially expressed genes correlated with a cancerous state of a mammalian cell, the method comprising the step of detecting at least one differentially expressed gene product in a test sample derived from a cell suspected of being cancerous, where the gene product comprises an amino acid sequence selected from the group consisting of SEQ ID NOS:1486-1542. Detection of the differentially expressed gene product is correlated with a cancerous state of the cell from which the test sample was derived.

In another aspect, the invention provides a library of polynucleotides, wherein at least one of the polynucleotides comprises the sequence information of a polynucleotide of the invention. In specific embodiments, the library is provided on a nucleic acid array. In some embodiments, the library is provided in a computer-readable format.

In another aspect, the invention provides a method of inhibiting tumor growth by modulating expression of a gene product, the gene product being encoded by a gene identified by a sequence selected from the group consisting of SEQ ID NOS:1-1485.

In another aspect, the invention provides a method of inhibiting tumor growth by modulating expression of a gene product, the gene product comprising an amino acid sequence selected from the group consisting of SEQ ID NOS: 1486-1542.

These and other objects, advantages, and features of the invention will become apparent to those persons skilled in the art upon reading the details of the invention as more fully described below.

## Detailed Description of the Invention

Before the present invention is described, it is to be understood that this invention is not limited to particular embodiments described, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting.

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are now described.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference. The citation of any publication is for its disclosure prior to the filing date and should not be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention.

It must be noted that as used herein and in the appended claims, the singular forms "a," "and," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a polynucleotide" includes a plurality of such polynucleotides and reference to "the colon cancer cell" includes reference to one or more cells and equivalents thereof known to those skilled in the art, and so forth.

The publications and applications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention. Further, the dates of publication provided may be different from the actual publication dates which may need to be independently confirmed.

Definitions

The terms "polynucleotide" and "nucleic acid," used interchangeably herein, refer to a polymeric forms of nucleotides of any length, either ribonucleotides or deoxynucleotides. Thus, these terms include, but are not limited to, single-, double-, or multi-stranded DNA or RNA, genomic DNA, cDNA, DNA-RNA hybrids, branched nucleic acid (see, *e.g.*, U.S. Pat. Nos. 5,124,246; 5,710,264; and 5,849,481), or a polymer comprising purine and pyrimidine bases or other natural, chemically or biochemically modified, non-natural, or derivatized nucleotide bases. These terms furhter include, but are not limited to, mRNA or cDNA that comprise intronic sequences (see, *e.g.*, Niwa et al. (1999) Cell 99(7):691-702). The backbone of the polynucleotide can comprise sugars and phosphate groups (as may typically be found in RNA or DNA), or modified or substituted sugar or phosphate groups. Alternatively, the backbone of the polynucleotide can comprise a polymer of synthetic subunits such as phosphoramidites and thus can be an oligodeoxynucleoside phosphoramidate or a mixed phosphoramidate-phosphodiester oligomer. Peyrottes et al. (1996) Nucl. Acids Res. 24:1841-1848; Chaturvedi et al. (1996) Nucl. Acids Res. 24:2318-2323. A polynuclotide may comprise modified nucleotides, such as methylated nucleotides and nucleotide analogs, uracyl, other sugars, and linking groups such as fluororibose and thioate, and nucleotide branches. The sequence of nucleotides may be interrupted by non-nucleotide components. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling component. Other types of modifications included in this definition are caps, substitution of one or more of the naturally occurring nucleotides with an analog, and introduction of means for attaching the polynucleotide to proteins, metal ions, labeling components, other polynucleotides, or a solid support.

The terms "polypeptide" and "protein," used interchangebly herein, refer to a polymeric form of amino acids of any length, which can include coded and non-coded amino acids, chemically or biochemically modified or derivatized amino acids, and polypeptides having modified peptide backbones. The term includes fusion proteins, including, but not limited to, fusion proteins with a heterologous amino acid sequence, fusions with heterologous and homologous leader sequences, with or without N-terminal methionine residues; immunologically tagged proteins; and the like.

"Diagnosis" as used herein generally includes determination of a subject's susceptibility to a disease or disorder, determination as to whether a subject is presently affected by a disease or disorder, prognosis of a subject affected by a disease or disorder (*e.g.*, identification of pre-metastatic or metastatic cancerous states, stages of cancer, or responsiveness of cancer to therapy), and therametrics (*e.g.*, monitoring a subject's condition to provide information as to the effect or efficacy of therapy).

"Sample" or "biological sample" as used herein encompasses a variety of sample types, and are generally meant to refer to samples of biological fluids or tissues, particularly samples obtained from tissues, especially from cells of the type associated with a disease or condition for which a diagnostic application is designed (e.g., ductal adenocarcinoma), and the like. "Sample" or "biological sample" are meant to encompass blood and other liquid samples of biological origin, solid tissue samples, such as a biopsy specimen or tissue cultures or cells derived therefrom and the progeny thereof. These terms encompass samples that have been manipulated in any way after their procurement as well as derivatives and fractions of samples, where the samples may be maniuplated by, for example, treatment with reagents, solubilization, or enrichment for certain components. The terms also encompass clinical samples, and also includes cells in cell culture, cell supernatants, cell lysates, serum, plasma, biological fluids, and tissue samples. Where the sample is solid tissue, the cells of the tissue can be dissociated or tissue sections can be analyzed.

The terms "treatment," "treating," "treat" and the like are used herein to generally refer to obtaining a desired pharmacologic and/or physiologic effect. The effect may be prophylactic in terms of completely or partially preventing a disease or symptom thereof and/or may be therapeutic in terms of a partial or complete stabilization or cure for a disease and/or adverse effect attributable to the disease. "Treatment" as used herein covers any treatment of a disease in a mammal, particularly a human, and includes: (a) preventing the disease or symptom from occurring in a subject which may be predisposed to the disease or symptom but has not yet been diagnosed as having it; (b) inhibiting the disease symptom, i.e., arresting its development; or relieving the disease symptom, i.e., causing regression of the disease or symptom.

The terms "individual," "subject," "host," and "patient," used interchangeably herein and refer to any mammalian subject for whom diagnosis, treatment, or therapy is desired, particularly humans. Other subjects may include cattle, dogs, cats, guinea pigs, rabbits, rats, mice, horses, and so on.

As used herein the term "isolated" refers to a polynucleotide, a polypeptide, an antibody, or a host cell that is in an environment different from that in which the polynucleotide, the polypeptide, the antibody, or the host cell naturally occurs. A polynucleotide, a polypeptide, an antibody, or a host cell which is isolated is generally substantially purified. As used herein, the term "substantially purified" refers to a compound (e.g., either a polynucleotide or a polypeptide or an antibody) that is removed from its natural environment and is at least 60% free, preferably 75% free, and most preferably 90% free from other components with which it is naturally associated. Thus, for example, a composition containing A is "substantially free of" B when at least 85% by weight of the total A+B in the composition is A. Preferably, A comprises at least about 90% by weight of the total of A+B in the composition, more preferably at least about 95% or even 99% by weight.

A "host cell," as used herein, refers to a microorganism or a eukaryotic cell or cell line cultured as a unicellular entity which can be, or has been, used as a recipient for a recombinant vector or other transfer polynucleotides, and include the progeny of the original cell which has been transfected. It is understood that the progeny of a single cell may not necessarily be completely identical in morphology or in genomic or total DNA complement as the original parent, due to natural, accidental, or deliberate mutation.

The terms "cancer," "neoplasm," "tumor," and "carcinoma," are used interchangeably herein to refer to cells which exhibit relatively autonomous growth, so that they exhibit an aberrant growth phenotype characterized by a significant loss of control of cell proliferation. In general, cells of interest for detection or treatment in the present application include precancerous (*e.g.*, benign), malignant, metastatic, and non-metastatic cells. Detection of cancerous cell is of particular interest.

The use of "e", as in 10e-3, indicates that the number to the left of "e" is raised to the power of the number to the right of "e" (thus, 10e-3 is $10^{-3}$).

The term "heterologous" as used herein in the context of, for example, heterologous nucleic acid or amino acid sequences, heterologous polypeptides, or heterologous nucleic acid, is meant to refer to material that originates from a source different from that with which it is joined or associated. For example, two DNA sequences are heterologous to one another if the sequences are from different genes or from different species. A recombinant host cell containing a sequence that is heterologous to the host cell can be, for example, a bacterial cell containing a sequence encoding a human polypeptide.

The invention relates to polynucleotides comprising the disclosed nucleotide sequences, to full length cDNA, mRNA, genomic sequences, and genes corresponding to these sequences and degenerate variants thereof, and to polypeptides encoded by the polynucleotides of the invention and polypeptide variants. The following detailed description describes the polynucleotide compositions encompassed by the invention, methods for obtaining cDNA or genomic DNA encoding a full-length gene product, expression of these polynucleotides and genes, identification of structural motifs of the polynucleotides and genes, identification of the function of a gene product encoded by a gene corresponding to a polynucleotide of the invention, use of the provided polynucleotides as probes and in mapping and in tissue profiling, use of the corresponding polypeptides and other gene products to raise antibodies, and use of the polynucleotides and their encoded gene products for therapeutic and diagnostic purposes.

Polynucleotide Compositions

The present invention provides isolated polynucleotides that represent genes that are differentially expressed in human cancer cells. The polynucleotides, as well as polypeptides encoded thereby, find use in a variety of therapeutic and diagnostic methods.

The scope of the invention with respect to compositions containing the isolated polynucleotides useful in the methods described herein includes, but is not necessarily limited to, polynucleotides having a sequence set forth in any one of the polynucleotide sequences provided herein; polynucleotides obtained from the biological materials described herein or other biological sources (particularly human sources) by hybridization under stringent conditions (particularly conditions of high stringency); genes corresponding to the provided polynucleotides; cDNAs corresponding to the provided polynucleotides; variants of the provided polynucleotides and their corresponding genes, particularly those variants that retain a biological activity of the encoded gene product (*e.g.*, a biological activity ascribed to a gene product corresponding to the provided polynucleotides as a result of the assignment of the gene product to a protein family(ies) and/or identification of a functional domain present in the gene product). Other nucleic acid compositions contemplated by and within the scope of the present invention will be readily apparent to one of ordinary skill in the art when provided with the disclosure here. "Polynucleotide" and "nucleic acid" as used herein with reference to nucleic acids of the composition is not intended to be limiting as to the length or structure of the nucleic acid unless specifically indicated.

The invention features polynucleotides that represent genes that are expressed in human tissue, specifically human breast tissue, particularly polynucleotides that are differentially expressed in cancerous breast cells. Nucleic acid compositions described herein of particular interest are at least about 15 bp in length, at least about 30 bp in length, at least about 50 bp in length, at least about 100 bp, at least about 200 bp in length, at least about 300 bp in length, at least about 500 bp in length, at least about 800 bp in length, at least about 1 kb in length, at least about 2.0 kb in length, at least about 3.0 kb in length, at least about 5 kb in length, at least about 10 kb in length, at least about 50kb in length and are usually less than about 200 kb in length. These polynucleotides (or polynucleotide fragments) have uses that include, but are not limited to, diagnostic probes and primers as starting materials for probes and primers, as discussed herein.

The subject polynucleotides usually comprise a sequence set forth in any one of the polynucleotide sequences provided herein, for example, in the sequence listing, incorporated by reference in a table (e.g. by an NCBI accession number), a cDNA deposited at the A.T.C.C., or a fragment or variant thereof. A "fragment" or "portion" of a polynucleotide is a contiguous sequence of residues at least about 10 nt to about 12 nt, 15 nt, 16 nt, 18 nt or 20 nt in length, usually at least about 22 nt, 24 nt, 25 nt, 30 nt, 40 nt, 50 nt, 60nt, 70 nt, 80 nt, 90 nt, 100 nt toat least about 150 nt, 200 nt, 250 nt, 300 nt, 350 nt, 400 nt, 500 nt, 800 nt or up to about 1000 nt, 1500 or 2000 nt in length. In some embodiments, a fragment of a polynucleotide is the coding sequence of a polynucleotide. A fragment of a polynucleotide may start at position 1 (i.e. the first nucleotide) of a nucleotide sequence provided herein, or may start at about position 10, 20, 30, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000, 1500 or 2000, or an ATG translational initiation codon of a nucleotide sequence provided herein. In this context "about" includes the particularly recited value or a value larger or smaller by several (5, 4, 3, 2, or 1) nucleotides. The described polynucleotides and fragments thereof find use as hybridization probes, PCR primers, BLAST probes, or as an identifying sequence, for example.

The subject nucleic acids may be variants or degenerate variants of a sequence provided herein. In general, a variants of a polynucleotide provided herein have a fragment of sequence identity that is greater than at least about 65%, greater than at least about 70%, greater than at least about 75%, greater than at least about 80%, greater than at least about 85%, or greater than at least about 90%, 95%, 96%, 97%, 98%, 99% or more (i.e. 100%) as compared to an identically sized fragment of a provided sequence. as determined by the Smith-Waterman homology search algorithm as implemented in MPSRCH program (Oxford Molecular). For the purposes of this invention, a preferred method of calculating percent identity is the Smith-Waterman algorithm. Global DNA sequence identity should be greater than 65% as determined by the Smith-Waterman homology search algorithm as implemented in MPSRCH program (Oxford Molecular) using an gap search with the following search parameters: gap open penalty, 12; and gap extension penalty, 1.

The subject nucleic acid compositions include full-length cDNAs or mRNAs that encompass an identifying sequence of contiguous nucleotides from any one of the polynucleotide sequences provided herein.

As discussed above, the polynucleotides useful in the methods described herein also include polynucleotide variants having sequence similarity or sequence identity. Nucleic acids having sequence similarity are detected by hybridization under low stringency conditions, for example, at 50°C and 10XSSC (0.9 M saline/0.09 M sodium citrate) and remain bound when subjected to washing at 55°C in 1XSSC. Sequence identity can be determined by hybridization under high stringency conditions, for example, at 50°C or higher and 0.1XSSC (9 mM saline/0.9 mM sodium citrate). Hybridization methods and conditions are well known in the art, see, *e.g.*, USPN 5,707,829. Nucleic acids that are substantially identical to the provided polynucleotide sequences, *e.g.* allelic variants, genetically altered versions of the gene, *etc.*, bind to the provided polynucleotide sequences under stringent hybridization conditions. By using probes, particularly labeled probes of DNA sequences, one can isolate homologous or related genes. The source of homologous genes can be any species, *e.g.* primate species, particularly human; rodents, such as rats and mice; canines, felines, bovines, ovines, equines, yeast, nematodes, *etc.*

In one embodiment, hybridization is performed using a fragment of at least 15 contiguous nucleotides (nt) of at least one of the polynucleotide sequences provided herein. That is, when at least 15 contiguous nt of one of the disclosed polynucleotide sequences is used as a probe, the probe will preferentially hybridize with a nucleic acid comprising the complementary sequence, allowing the identification and retrieval of the nucleic acids that uniquely hybridize to the selected probe. Probes from more than one polynucleotide sequence provided herein can hybridize with the same nucleic acid if the cDNA from which they were derived corresponds to one mRNA.

Polynucleotides contemplated for use in the invention also include those having a sequence of naturally occurring variants of the nucleotide sequences (*e.g.*, degenerate variants (*e.g.*, sequences that encode the same polypeptides but, due to the degenerate nature of the genetic code, different in nucleotide sequence), allelic variants, *etc.*). Variants of the polynucleotides contemplated by the invention are identified by hybridization of putative variants with nucleotide sequences disclosed herein, preferably by hybridization under stringent conditions. For example, by using appropriate wash conditions, variants of the polynucleotides described herein can be identified where the allelic variant exhibits at most about 25-30% base pair (bp) mismatches relative to the selected polynucleotide probe. In general, allelic variants contain 15-25% bp mismatches, and can contain as little as even 5-15%, or 2-5%, or 1-2% bp mismatches, as well as a single bp mismatch.

The invention also encompasses homologs corresponding to any one of the polynucleotide sequences provided herein, where the source of homologous genes can be any mammalian species, *e.g.*, primate species, particularly human; rodents, such as rats; canines, felines, bovines, ovines, equines, yeast, nematodes, etc. Between mammalian species, *e.g.*, human and mouse, homologs generally have substantial sequence similarity, *e.g.*, at least 75% sequence identity, usually at least 80%%, at least 85, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or even 100% identity between nucleotide sequences. Sequence similarity is calculated based on a reference sequence, which may be a subset of a larger sequence, such as a conserved motif, coding region, flanking region, *etc.* A reference sequence will usually be at least about a fragment of a polynucleotide sequence and may extend to the complete sequence that is being compared. Algorithms for sequence analysis are known in the art, such as gapped BLAST, described in Altschul, et al. *Nucleic Acids Res.* (1997) *25*:3389-3402, or TeraBLAST available from TimeLogic Corp. (Crystal Bay, Nevada).

Moreover, representative examples of polynucleotide fragments of the invention (useful, for example, as probes), include, for example, fragments comprising, or alternatively consisting of, a sequence from about nucleotide number 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-450, 451-500, 501-550, 551-600, 651-700,701- 750, 751-800, 800-850, 851-900, 901-950,951-1000, 1001-1050, 1051-1100, 1101-1150, 1151-1200, 1201-1250, 1251-1300, 1301-1350, 1351-1400, 1401-1450, 1451-1500, 1501-1550, 1551-1600, 1601-1650, 1651-1700, 1701-1750, 1751-1800, 1801-1850, 1851-1900, 1901-1950, 1951-2000, 2001-2050, 2051-2100, 2101-2150, 2151-2200, 2201-2250, 2251-2300, 2301-2350, 2351-2400, 2401-2450, 2451-2500, 2501-2550, 2551-2600, 2601-2650, 2651-2700, 2701-2750, 2751-2800, 2801-2850, 2851-2900, 2901-2950, 2951-3000, 3001-3050, 3051-3100, 3101-3150, 3151-3200, 3201-3250, 3251-3300, 3301-3350, 3351-3400, 3401-3450, 3451-3500, 3501-3550, 3551-3600, 3601-3650, 3651-3700, 3701-3750, 3751-3800, 3801-3850, 3851-3900, 3901-3950, 3951-4000, 4001-4050, 4051-4100, 4101-4150, 4151-4200, 4201-4250, 4251-4300, 4301-4350, 4351-4400, 4401-4450, 4451-4500, 4501-4550, 4551-4600, 4601-4650, 4651-4700, 4701-4750, 4751-4800, 4801-4850, 4851-4900, 4901-4950, 4951-5000, 5001-5050, 5051-5100, 5101-5150, 5151-5200, 5201-5250, 5251-5300, 5301-5350, 5351-5400, 5401- 5450, 5451-5500, 5501-5550, 5551-5600, 5601-5650, 5651-5700, 5701-5750, 5751-5800, 5801-5850, 5851-5900, 5901-5950, 5951-6000, 6001-6050, 6051-6100, 6101-6150, and 6151 of a subject nucleic acid, or the complementary strand thereto. In this context "about" includes the particularly recited range or a range larger or smaller by several (5, 4, 3, 2, or 1) nucleotides, at either terminus or at both termini. In some embodiments, these fragments encode a polypeptide which has a functional activity (e.g., biological activity) whereas in other embodiments, these fragments are probes, or starting materials for probes. Polynucleotides which hybridize to one or more of these nucleic acid molecules under stringent hybridization conditions or alternatively, under lower stringency conditions, are also encompassed by the invention, as are polypeptides encoded by these polynucleotides or fragments.

The subject nucleic acids can be cDNAs or genomic DNAs, as well as fragments thereof, particularly fragments that encode a biologically active gene product and/or are useful in the methods disclosed herein (e.g., in diagnosis, as a unique identifier of a differentially expressed gene of interest, etc.). The term "cDNA" as used herein is intended to include all nucleic acids that share the arrangement of sequence elements found in native mature mRNA species, where sequence elements are exons and 3' and 5' non-coding regions. Normally mRNA species have contiguous exons, with the intervening introns, when present, being removed by nuclear RNA splicing, to create a continuous open reading frame encoding a polypeptide. mRNA species can also exist with both exons and introns, where the introns may be removed by alternative splicing. Furthermore it should be noted that different species of mRNAs encoded by the same genomic sequence can exist at varying levels in a cell, and detection of these various levels of mRNA species can be indicative of differential expression of the encoded gene product in the cell.

A genomic sequence of interest comprises the nucleic acid present between the initiation codon and the stop codon, as defined in the listed sequences, including all of the introns that are normally present in a native chromosome. It can further include the 3' and 5' untranslated regions found in the mature mRNA. It can further include specific transcriptional and translational regulatory sequences, such as promoters, enhancers, *etc.*, including about 1 kb, but possibly more, of flanking genomic DNA at either the 5' and 3' end of the transcribed region. The genomic DNA can be isolated as a fragment of 100 kbp or smaller; and substantially free of flanking chromosomal sequence. The genomic DNA flanking the coding region, either 3' and 5', or internal regulatory sequences as sometimes found in introns, contains sequences required for proper tissue, stage-specific, or disease-state specific expression.

The nucleic acid compositions of the subject invention can encode all or a part of the naturally-occurring polypeptides. Double or single stranded fragments can be obtained from the DNA sequence by chemically synthesizing oligonucleotides in accordance with conventional methods, by restriction enzyme digestion, by PCR amplification, *etc.*

Probes specific to the polynucleotides described herein can be generated using the polynucleotide sequences disclosed herein. The probes are usually a fragment of a polynucleotide sequences provided herein. The probes can be synthesized chemically or can be generated from longer polynucleotides using restriction enzymes. The probes can be labeled, for example, with a radioactive, biotinylated, or fluorescent tag. Preferably, probes are designed based upon an identifying sequence of any one of the polynucleotide sequences provided herein. More preferably, probes are designed based on a contiguous sequence of one of the subject polynucleotides that remain unmasked following application of a masking program for masking low complexity (*e.g.*, XBLAST, RepeatMasker, etc.) to the sequence., *i.e.*, one would select an unmasked region, as indicated by the polynucleotides outside the poly-n stretches of the masked sequence produced by the masking program.

The polynucleotides of interest in the subject invention are isolated and obtained in substantial purity, generally as other than an intact chromosome. Usually, the polynucleotides, either as DNA or RNA, will be obtained substantially free of other naturally-occurring nucleic acid sequences that they are usually associated with , generally being at least about 50%, usually at least about 90% pure and are typically "recombinant", *e.g.*, flanked by one or more nucleotides with which it is not normally associated on a naturally occurring chromosome.

The polynucleotides described herein can be provided as a linear molecule or within a circular molecule, and can be provided within autonomously replicating molecules (vectors) or within molecules without replication sequences. Expression of the polynucleotides can be regulated by their own or by other regulatory sequences known in the art. The polynucleotides can be introduced into suitable host cells using a variety of techniques available in the art, such as transferrin polycation-mediated DNA transfer, transfection with naked or encapsulated nucleic acids, liposome-mediated DNA transfer, intracellular transportation of DNA-coated latex beads, protoplast fusion, viral infection, electroporation, gene gun, calcium phosphate-mediated transfection, and the like.

The nucleic acid compositions described herein can be used to, for example, produce polypeptides, as probes for the detection of mRNA in biological samples (*e.g.*, extracts of human cells) or cDNA produced from such samples, to generate additional copies of the polynucleotides, to generate ribozymes or antisense oligonucleotides, and as single stranded DNA probes or as triple-strand forming oligonucleotides. The probes described herein can be used to, for example, determine the presence or absence of any one of the polynucleotide provided herein or variants thereof in a sample. These and other uses are described in more detail below.The subject nucleic acid compositions can be used, for example, to produce polypeptides, as probes for the detection of mRNA of the invention in biological samples (e.g., extracts of human cells) to generate additional copies of the polynucleotides, to generate ribozymes or antisense oligonucleotides, and as single stranded DNA probes or as triple-strand forming oligonucleotides. The probes described herein can be used to, for example, determine the presence or absence of the polynucleotide sequences as shown in SEQ ID NOS:1-1485 or variants thereof in a sample. These and other uses are described in more detail below.

Use of Polynucleotides to Obtain Full-Length cDNA, Gene, and Promoter Region

In one embodiment, the polynucleotides are useful as starting materials to construct larger molecules. In one example, the polynucleotides of the invention are used to construct polynucleotides that encode a larger polypeptide (*e.g.*, up to the full-length native polypeptide as well as fusion proteins comprising all or a portion of the native polypeptide) or may be used to produce haptens of the polypeptide (*e.g.*, polypeptides useful to generate antibodies).

In one particular example, the polynucleotides of the invention are used to make or isolate cDNA molecules encoding all or portion of a naturally-occuring polypeptide. Full-length cDNA molecules comprising the disclosed polynucleotides are obtained as follows. A polynucleotide having a sequence of one of SEQ ID NOS:1-1485, or a portion thereof comprising at least 12, 15, 18, or 20 nt, is used as a hybridization probe to detect hybridizing members of a cDNA library using probe design methods, cloning methods, and clone selection techniques such as those described in USPN 5,654,173. Libraries of cDNA are made from selected tissues, such as normal or tumor tissue, or from tissues of a mammal treated with, for example, a pharmaceutical agent. Preferably, the tissue is the same as the tissue from which the polynucleotides of the invention were isolated, as both the polynucleotides described herein and the cDNA represent expressed genes. Most preferably, the cDNA library is made from the biological material described herein in the Examples. The choice of cell type for library construction can be made after the identity of the protein encoded by the gene corresponding to the polynucleotide of the invention is known. This will indicate which tissue and cell types are likely to express the related gene, and thus represent a suitable source for the mRNA for generating the cDNA. Where the provided polynucleotides are isolated from cDNA libraries, the libraries are prepared from mRNA of human prostate cells, more preferably, human prostate cancer cells

Techniques for producing and probing nucleic acid sequence libraries are described, for example, in Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Ed., (1989) Cold Spring

Harbor Press, Cold Spring Harbor, NY. The cDNA can be prepared by using primers based on polynucleotides comprising a sequence of SEQ ID NOS:1-1485. In one embodiment, the cDNA library can be made from only poly-adenylated mRNA. Thus, poly-T primers can be used to prepare cDNA from the mRNA.

Members of the library that are larger than the provided polynucleotides, and preferably that encompass the complete coding sequence of the native message, are obtained. In order to confirm that the entire cDNA has been obtained, RNA protection experiments are performed as follows. Hybridization of a full-length cDNA to an mRNA will protect the RNA from RNase degradation. If the cDNA is not full length, then the portions of the mRNA that are not hybridized will be subject to RNase degradation. This is assayed, as is known in the art, by changes in electrophoretic mobility on polyacrylamide gels, or by detection of released monoribonucleotides. Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Ed., (1989) Cold Spring Harbor Press, Cold Spring Harbor, NY. In order to obtain additional sequences 5' to the end of a partial cDNA, 5' RACE (PCR Protocols: A Guide to Methods and Applications, (1990) Academic Press, Inc.) can be performed.

Genomic DNA is isolated using the provided polynucleotides in a manner similar to the isolation of full-length cDNAs. Briefly, the provided polynucleotides, or portions thereof, are used as probes to libraries of genomic DNA. Preferably, the library is obtained from the cell type that was used to generate the polynucleotides of the invention, but this is not essential. Most preferably, the genomic DNA is obtained from the biological material described herein in the Examples. Such libraries can be in vectors suitable for carrying large segments of a genome, such as P1 or YAC, as described in detail in Sambrook et al., supra, 9.4-9.30. In addition, genomic sequences can be isolated from human BAC libraries, which are commercially available from Research Genetics, Inc., Huntsville, Alabama, USA, for example. In order to obtain additional 5' or 3' sequences, chromosome walking is performed, as described in Sambrook et al., such that adjacent and overlapping fragments of genomic DNA are isolated. These are mapped and pieced together, as is known in the art, using restriction digestion enzymes and DNA ligase.

Using the polynucleotide sequences of the invention, corresponding full-length genes can be isolated using both classical and PCR methods to construct and probe cDNA libraries. Using either method, Northern blots, preferably, are performed on a number of cell types to determine which cell lines express the gene of interest at the highest level. Classical methods of constructing cDNA libraries are taught in Sambrook et al., supra. With these methods, cDNA can be produced from mRNA and inserted into viral or expression vectors. Typically, libraries of mRNA comprising poly(A) tails can be produced with poly(T) primers. Similarly, cDNA libraries can be produced using the instant sequences as primers.

PCR methods are used to amplify the members of a cDNA library that comprise the desired insert. In this case, the desired insert will contain sequence from the full length cDNA that corresponds to the instant polynucleotides. Such PCR methods include gene trapping and RACE methods. Gene trapping entails inserting a member of a cDNA library into a vector. The vector then is denatured to produce single stranded molecules. Next, a substrate-bound probe, such as a biotinylated oligo, is used to trap cDNA inserts of interest. Biotinylated probes can be linked to an avidin-bound solid substrate. PCR methods can be used to amplify the trapped cDNA. To trap sequences corresponding to the full length genes, the labeled probe sequence is based on the polynucleotide sequences of the invention. Random primers or primers specific to the library vector can be used to amplify the trapped cDNA. Such gene trapping techniques are described in Gruber et al., WO 95/04745 and Gruber et al., USPN 5,500,356. Kits are commercially available to perform gene trapping experiments from, for example, Life Technologies, Gaithersburg, Maryland, USA.

"Rapid amplification of cDNA ends," or RACE, is a PCR method of amplifying cDNAs from a number of different RNAs. The cDNAs are ligated to an oligonucleotide linker, and amplified by PCR using two primers. One primer is based on sequence from the instant polynucleotides, for which full length sequence is desired, and a second primer comprises sequence that hybridizes to the oligonucleotide linker to amplify the cDNA. A description of this method is reported in WO 97/19110. In preferred embodiments of RACE, a common primer is designed to anneal to an arbitrary adaptor sequence ligated to cDNA ends (Apte and Siebert, Biotechniques (1993) 15:890-893; Edwards et al., Nuc. Acids Res. (1991) 19:5227-5232). When a single gene-specific RACE primer is paired with the common primer, preferential amplification of sequences between the single gene specific primer and the common primer occurs. Commercial cDNA pools modified for use in RACE are available.

Another PCR-based method generates full-length cDNA library with anchored ends without needing specific knowledge of the cDNA sequence. The method uses lock-docking primers (I-VI), where one primer, poly TV (I-III) locks over the polyA tail of eukaryotic mRNA producing first strand synthesis and a second primer, polyGH (IV-VI) locks onto the polyC tail added by terminal deoxynucleotidyl transferase (TdT)(see, e.g., WO 96/40998).

The promoter region of a gene generally is located 5' to the initiation site for RNA polymerase II. Hundreds of promoter regions contain the "TATA" box, a sequence such as TATTA or TATAA, which is sensitive to mutations. The promoter region can be obtained by performing 5' RACE using a primer from the coding region of the gene. Alternatively, the cDNA can be used as a probe for the genomic sequence, and the region 5' to the coding region is identified by "walking up." If the gene is highly expressed or differentially expressed, the promoter from the gene can be of use in a regulatory construct for a heterologous gene.

Once the full-length cDNA or gene is obtained, DNA encoding variants can be prepared by site-directed mutagenesis, described in detail in Sambrook et al., 15.3-15.63. The choice of codon or nucleotide to be replaced can be based on disclosure herein on optional changes in amino acids to achieve altered protein structure and/or function.

As an alternative method to obtaining DNA or RNA from a biological material, nucleic acid comprising nucleotides having the sequence of one or more polynucleotides of the invention can be synthesized. Thus, the invention encompasses nucleic acid molecules ranging in length from 15 nt (corresponding to at least 15 contiguous nt of one of SEQ ID NOS:1-1485) up to a maximum length suitable for one or more biological manipulations, including replication and expression, of the nucleic acid molecule. The invention includes but is not limited to (a) nucleic acid having the size of a full gene, and comprising at least one of SEQ ID NOS:1-1485; (b) the nucleic acid of (a) also comprising at least one additional gene, operably linked to permit expression of a fusion protein; (c) an expression vector comprising (a) or (b); (d) a plasmid comprising (a) or (b); and (e) a recombinant viral particle comprising (a) or (b). Once provided with the polynucleotides disclosed herein, construction or preparation of (a) - (e) are well within the skill in the art.

The sequence of a nucleic acid comprising at least 15 contiguous nt of at least any one of SEQ ID NOS:1-1485, preferably the entire sequence of at least any one of SEQ ID NOS:1-1485, is not limited and can be any sequence of A, T, G, and/or C (for DNA) and A, U, G, and/or C (for RNA) or modified bases thereof, including inosine and pseudouridine. The choice of sequence will depend on the desired function and can be dictated by coding regions desired, the intron-like regions desired, and the regulatory regions desired. Where the entire sequence of any one of SEQ ID NOS:1-1485 is within the nucleic acid, the nucleic acid obtained is referred to herein as a polynucleotide comprising the sequence of any one of SEQ ID NOS:1-1485.

Expression of Polypeptide Encoded by Full-Length cDNA or Full-Length Gene

The provided polynucleotides (*e.g.*, a polynucleotide having a sequence of one of SEQ ID NOS:1-1485), the corresponding cDNA, or the full-length gene is used to express a partial or complete gene product. Constructs of polynucleotides having sequences of SEQ ID NOS:1-1485 can also be generated synthetically. Alternatively, single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides is described by, e.g., Stemmer et al., Gene (Amsterdam) (1995) 164(1):49-53. In this method, assembly PCR (the synthesis of long DNA sequences from large numbers of oligodeoxyribonucleotides (oligos)) is described. The method is derived from DNA shuffling (Stemmer, Nature (1994) 370:389-391), and does not rely on DNA ligase, but instead relies on DNA polymerase to build increasingly longer DNA fragments during the assembly process.

Appropriate polynucleotide constructs are purified using standard recombinant DNA techniques as described in, for example, Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual, 2nd Ed.*, (1989) Cold Spring Harbor Press, Cold Spring Harbor, NY, and under current regulations described in United States Dept. of HHS, National Institute of Health (NIH) Guidelines for Recombinant DNA Research. The gene product encoded by a polynucleotide of the invention is expressed in any expression system, including, for example, bacterial, yeast, insect, amphibian and mammalian systems. Vectors, host cells and methods for obtaining expression in same are well known in the art. Suitable vectors and host cells are described in USPN 5,654,173.

Polynucleotide molecules comprising a polynucleotide sequence provided herein are generally propagated by placing the molecule in a vector. Viral and non-viral vectors are used, including plasmids. The choice of plasmid will depend on the type of cell in which propagation is desired and the purpose of propagation. Certain vectors are useful for amplifying and making large amounts of the desired DNA sequence. Other vectors are suitable for expression in cells in culture. Still other vectors are suitable for transfer and expression in cells in a whole animal or person. The choice of appropriate vector is well within the skill of the art. Many such vectors are available commercially. Methods for preparation of vectors comprising a desired sequence are well known in the art.

The polynucleotides set forth in SEQ ID NOS:1-1485 or their corresponding full-length polynucleotides are linked to regulatory sequences as appropriate to obtain the desired expression properties. These can include promoters (attached either at the 5' end of the sense strand or at the 3' end of the antisense strand), enhancers, terminators, operators, repressors, and inducers. The promoters can be regulated or constitutive. In some situations it may be desirable to use conditionally active promoters, such as tissue-specific or developmental stage-specific promoters. These are linked to the desired nucleotide sequence using the techniques described above for linkage to vectors. Any techniques known in the art can be used.

When any of the above host cells, or other appropriate host cells or organisms, are used to replicate and/or express the polynucleotides or nucleic acids of the invention, the resulting replicated nucleic acid, RNA, expressed protein or polypeptide, is within the scope of the invention as a product of the host cell or organism. The product is recovered by any appropriate means known in the art.

Once the gene corresponding to a selected polynucleotide is identified, its expression can be regulated in the cell to which the gene is native. For example, an endogenous gene of a cell can be regulated by an exogenous regulatory sequence as disclosed in USPN 5,641,670.

Identification of Functional and Structural Motifs

Translations of the nucleotide sequence of the provided polynucleotides, cDNAs or full genes can be aligned with individual known sequences. Similarity with individual sequences can be used to determine the activity of the polypeptides encoded by the polynucleotides of the invention. Also, sequences exhibiting similarity with more than one individual sequence can exhibit activities that are characteristic of either or both individual sequences.

The full length sequences and fragments of the polynucleotide sequences of the nearest neighbors as identified through, for example, BLAST-based searching,can be used as probes and primers to identify and isolate the full length sequence corresponding to provided polynucleotides.

The nearest neighbors can indicate a tissue or cell type to be used to construct a library for the full-length sequences corresponding to the provided polynucleotides.

Typically, a selected polynucleotide is translated in all six frames to determine the best alignment with the individual sequences. The sequences disclosed herein in the Sequence Listing are in a 5' to 3' orientation and translation in three frames can be sufficient (with a few specific exceptions as described in the Examples). These amino acid sequences are referred to, generally, as query sequences, which will be aligned with the individual sequences. Databases with individual sequences are described in "Computer Methods for Macromolecular Sequence Analysis" *Methods in Enzymology* (1996) *266*, Doolittle, Academic Press, Inc., a division of Harcourt Brace & Co., San Diego, California, USA. Databases include GenBank, EMBL, and DNA Database of Japan (DDBJ).

Query and individual sequences can be aligned using the methods and computer programs described above, and include BLAST 2.0, available over the world wide web at a site supported by the National Center for Biotechnology Information, which is supported by the National Library of Medicine and the National Institutes of Health, or TeraBLAST available from TimeLogic Corp. (Crystal Bay, Nevada). See also Altschul, et al. Nucleic Acids Res. (1997) 25:3389-3402. Another alignment algorithm is Fasta, available in the Genetics Computing Group (GCG) package, Madison, Wisconsin, USA, a wholly owned subsidiary of Oxford Molecular Group, Inc. Other techniques for alignment are described in Doolittle, supra. Preferably, an alignment program that permits gaps in the sequence is utilized to align the sequences. The Smith-Waterman is one type of algorithm that permits gaps in sequence alignments. See Meth. Mol. Biol. (1997) *70*: 173-187. Also, the GAP program using the Needleman and Wunsch alignment method can be utilized to align sequences. An alternative search strategy uses MPSRCH software, which runs on a MASPAR computer. MPSRCH uses a Smith-Waterman algorithm to score sequences on a massively parallel computer. This approach improves ability to identify sequences that are distantly related matches, and is especially tolerant of small gaps and nucleotide sequence errors. Amino acid sequences encoded by the provided polynucleotides can be used to search both protein and DNA databases. Incorporated herein by reference are all sequences that have been made public as of the filing date of this application by any of the DNA or protein sequence databases, including the patent databases (*e.g.*, GeneSeq). Also incorporated by reference are those sequences that have been submitted to these databases as of the filing date of the present application but not made public until after the filing date of the present application.

Results of individual and query sequence alignments can be divided into three categories: high similarity, weak similarity, and no similarity. Individual alignment results ranging from high similarity to weak similarity provide a basis for determining polypeptide activity and/or structure. Parameters for categorizing individual results include: percentage of the alignment region length where the strongest alignment is found, percent sequence identity, and p value. The percentage of the alignment region length is calculated by counting the number of residues of the individual sequence found in the region of strongest alignment, *e.g.*, contiguous region of the individual sequence that contains the greatest number of residues that are identical to the residues of the corresponding region of the aligned query sequence. This number is divided by the total residue length of the query sequence to calculate a percentage. For example, a query sequence of 20 amino acid residues might be aligned with a 20 amino acid region of an individual sequence. The individual sequence might be identical to amino acid residues 5, 9-15, and 17-19 of the query sequence. The region of strongest alignment is thus the region stretching from residue 9-19, an 11 amino acid stretch. The percentage of the alignment region length is: 11 (length of the region of strongest alignment) divided by (query sequence length) 20 or 55%.

Percent sequence identity is calculated by counting the number of amino acid matches between the query and individual sequence and dividing total number of matches by the number of residues of the individual sequences found in the region of strongest alignment. Thus, the percent identity in the example above would be 10 matches divided by 11 amino acids, or approximately, 90.9%

P value is the probability that the alignment was produced by chance. For a single alignment, the p value can be calculated according to Karlin et al., Proc. Natl. Acad. Sci. (1990) 87:2264 and Karlin et al., Proc. Natl. Acad. Sci. (1993) 90. The p value of multiple alignments using the same query sequence can be calculated using an heuristic approach described in Altschul et al., Nat. Genet. (1994) 6:119. Alignment programs, such as BLAST or TeraBLAST, can calculate the p value. See also Altschul et al., Nucleic Acids Res. (1997) 25:3389-3402.

Another factor to consider for determining identity or similarity is the location of the similarity or identity. Strong local alignment can indicate similarity even if the length of alignment is short. Sequence identity scattered throughout the length of the query sequence also can indicate a similarity between the query and profile sequences. The boundaries of the region where the sequences align can be determined according to Doolittle, supra; BLAST 2.0 (see, e.g., Altschul, et al. Nucleic Acids Res. (1997) 25:3389-3402), TeraBLAST (available from TimeLogic Corp., Crystal Bay, Nevada), or FAST programs; or by determining the area where sequence identity is highest.

High Similarity. In general, in alignment results considered to be of high similarity, the percent of the alignment region length is typically at least about 55% of total length query sequence; more typically, at least about 58%; even more typically; at least about 60% of the total residue length of the query sequence. Usually, percent length of the alignment region can be as much as about 62%; more usually, as much as about 64%; even more usually, as much as about 66%. Further, for high similarity, the region of alignment, typically, exhibits at least about 75% of sequence identity; more typically, at least about 78%; even more typically; at least about 80% sequence identity. Usually, percent sequence identity can be as much as about 82%; more usually, as much as about 84%; even more usually, as much as about 86%.

The p value is used in conjunction with these methods. If high similarity is found, the query sequence is considered to have high similarity with a profile sequence when the p value is less than or equal to about 10e-2; more usually; less than or equal to about 10e-3; even more usually; less than or equal to about 10e-4. More typically, the p value is no more than about 10e-5; more typically; no more than or equal to about 10e-10; even more typically, no more than or equal to about 10e-15 for the query sequence to be considered high similarity.

Weak Similarity. In general, where alignment results considered to be of weak similarity, there is no minimum percent length of the alignment region nor minimum length of alignment. A better showing of weak similarity is considered when the region of alignment is, typically, at least about 15 amino acid residues in length; more typically, at least about 20; even more typically, at least about 25 amino acid residues in length. Usually, length of the alignment region can be as much as about 30 amino acid residues; more usually, as much as about 40; even more usually, as much as about 60 amino acid residues. Further, for weak similarity, the region of alignment, typically, exhibits at least about 35% of sequence identity; more typically, at least about 40%; even more typically, at least about 45% sequence identity. Usually, percent sequence identity can be as much as about 50%; more usually, as much as about 55%; even more usually, as much as about 60%.

If low similarity is found, the query sequence is considered to have weak similarity with a profile sequence when the p value is usually less than or equal to about 10e-2; more usually, less than or equal to about 10e-3; even more usually; less than or equal to about 10e-4. More typically, the p value is no more than about 10e-5; more usually; no more than or equal to about 10e-10; even more usually, no more than or equal to about 10e-15 for the query sequence to be considered weak similarity.

Similarity Determined by Sequence Identity Alone. Sequence identity alone can be used to determine similarity of a query sequence to an individual sequence and can indicate the activity of the sequence. Such an alignment, preferably, permits gaps to align sequences. Typically, the query sequence is related to the profile sequence if the sequence identity over the entire query sequence is at least about 15%; more typically, at least about 20%; even more typically, at least about 25%; even more typically, at least about 50%. Sequence identity alone as a measure of similarity is most useful when the query sequence is usually, at least 80 residues in length; more usually, at least 90 residues in length; even more usually, at least 95 amino acid residues in length. More typically, similarity can be concluded based on sequence identity alone when the query sequence is preferably 100 residues in length; more preferably, 120 residues in length; even more preferably, 150 amino acid residues in length.

Alignments with Profile and Multiple Aligned Sequences. Translations of the provided polynucleotides can be aligned with amino acid profiles that define either protein families or common motifs. Also, translations of the provided polynucleotides can be aligned to multiple sequence alignments (MSA) comprising the polypeptide sequences of members of protein families or motifs. Similarity or identity with profile sequences or MSAs can be used to determine the activity of the gene products (e.g., polypeptides) encoded by the provided polynucleotides or corresponding cDNA or genes. For example, sequences that show an identity or similarity with a chemokine profile or MSA can exhibit chemokine activities.

Profiles can be designed manually by (1) creating an MSA, which is an alignment of the amino acid sequence of members that belong to the family and (2) constructing a statistical representation of the alignment. Such methods are described, for example, in Birney et al., Nucl. Acid Res. (1996) 24(14): 2730-2739. MSAs of some protein families and motifs are publicly available. For example, the Genome Sequencing Center at thw Washington University School of Medicine provides a web set (Pfam) which provides MSAs of 547 different families and motifs. These MSAs are described also in Sonnhammer et al., Proteins (1997) 28: 405-420. Other sources over the world wide web include the site supported by the European Molecular Biology Laboratories in Heidelberg, Germany. A brief description of these MSAs is reported in Pascarella et al., Prot. Eng. (1996) 9(3):249-251. Techniques for building profiles from MSAs are described in Sonnhammer et al., supra; Birney et al., supra; and "Computer Methods for Macromolecular Sequence Analysis," Methods in Enzymology (1996) 266, Doolittle, Academic Press, Inc., San Diego, California, USA.

Similarity between a query sequence and a protein family or motif can be determined by (a) comparing the query sequence against the profile and/or (b) aligning the query sequence with the members of the family or motif. Typically, a program such as Searchwise is used to compare the query sequence to the statistical representation of the multiple alignment, also known as a profile (see Birney et al., supra). Other techniques to compare the sequence and profile are described in Sonnhammer et al., supra and Doolittle, supra.

Next, methods described by Feng et al., J. Mol. Evol. (1987) 25:351 and Higgins et al., CABIOS (1989) 5:151 can be used align the query sequence with the members of a family or motif, also known as a MSA. Sequence alignments can be generated using any of a variety of software tools. Examples include PileUp, which creates a multiple sequence alignment, and is described in Feng et al., J. Mol. Evol. (1987) 25:351. Another method, GAP, uses the alignment method of Needleman et al., J. Mol. Biol. (1970) 48:443. GAP is best suited for global alignment of sequences. A third method, BestFit, functions by inserting gaps to maximize the number of matches using the local homology algorithm of Smith et al., Adv. Appl. Math. (1981) 2:482. In general, the following factors are used to determine if a similarity between a query sequence and a profile or MSA exists: (1)

number of conserved residues found in the query sequence, (2) percentage of conserved residues found in the query sequence, (3) number of frameshifts, and (4) spacing between conserved residues.

Some alignment programs that both translate and align sequences can make any number of frameshifts when translating the nucleotide sequence to produce the best alignment. The fewer frameshifts needed to produce an alignment, the stronger the similarity or identity between the query and profile or MSAs. For example, a weak similarity resulting from no frameshifts can be a better indication of activity or structure of a query sequence, than a strong similarity resulting from two frameshifts. Preferably, three or fewer frameshifts are found in an alignment; more preferably two or fewer frameshifts; even more preferably, one or fewer frameshifts; even more preferably, no frameshifts are found in an alignment of query and profile or MSAs.

Conserved residues are those amino acids found at a particular position in all or some of the family or motif members. Alternatively, a position is considered conserved if only a certain class of amino acids is found in a particular position in all or some of the family members. For example, the N-terminal position can contain a positively charged amino acid, such as lysine, arginine, or histidine.

Typically, a residue of a polypeptide is conserved when a class of amino acids or a single amino acid is found at a particular position in at least about 40% of all class members; more typically, at least about 50%; even more typically, at least about 60% of the members. Usually, a residue is conserved when a class or single amino acid is found in at least about 70% of the members of a family or motif; more usually, at least about 80%; even more usually, at least about 90%; even more usually, at least about 95%.

A residue is considered conserved when three unrelated amino acids are found at a particular position in some or all of the members; more usually, two unrelated amino acids. These residues are conserved when the unrelated amino acids are found at particular positions in at least about 40% of all class member; more typically, at least about 50%; even more typically, at least about 60% of the members. Usually, a residue is conserved when a class or single amino acid is found in at least about 70% of the members of a family or motif; more usually, at least about 80%; even more usually, at least about 90%; even more usually, at least about 95%.

A query sequence has similarity to a profile or MSA when the query sequence comprises at least about 25% of the conserved residues of the profile or MSA; more usually, at least about 30%; even more usually; at least about 40%. Typically, the query sequence has a stronger similarity to a profile sequence or MSA when the query sequence comprises at least about 45% of the conserved residues of the profile or MSA; more typically, at least about 50%; even more typically, at least about 55%.

<u>Identification of Secreted & Membrane-Bound Polypeptides.</u> Both secreted and membrane-bound polypeptides of the present invention are of particular interest. For example, levels of secreted polypeptides can be assayed in body fluids that are convenient, such as blood, plasma, serum, and

other body fluids such as urine, prostatic fluid and semen. Membrane-bound polypeptides are useful for constructing vaccine antigens or inducing an immune response. Such antigens would comprise all or part of the extracellular region of the membrane-bound polypeptides. Because both secreted and membrane-bound polypeptides comprise a fragment of contiguous hydrophobic amino acids, hydrophobicity predicting algorithms can be used to identify such polypeptides.

A signal sequence is usually encoded by both secreted and membrane-bound polypeptide genes to direct a polypeptide to the surface of the cell. The signal sequence usually comprises a stretch of hydrophobic residues. Such signal sequences can fold into helical structures. Membrane-bound polypeptides typically comprise at least one transmembrane region that possesses a stretch of hydrophobic amino acids that can transverse the membrane. Some transmembrane regions also exhibit a helical structure. Hydrophobic fragments within a polypeptide can be identified by using computer algorithms. Such algorithms include Hopp & Woods, Proc. Natl. Acad. Sci. USA (1981) 78:3824-3828; Kyte & Doolittle, J. Mol. Biol. (1982) 157: 105-132; and RAOAR algorithm, Degli Esposti et al., Eur. J. Biochem. (1990) 190: 207-219.

Another method of identifying secreted and membrane-bound polypeptides is to translate the polynucleotides of the invention in all six frames and determine if at least 8 contiguous hydrophobic amino acids are present. Those translated polypeptides with at least 8; more typically, 10; even more typically, 12 contiguous hydrophobic amino acids are considered to be either a putative secreted or membrane bound polypeptide. Hydrophobic amino acids include alanine, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, threonine, tryptophan, tyrosine, and valine

Identification of the Function of an Expression Product of a Full-Length Gene

Ribozymes, antisense constructs, and dominant negative mutants can be used to determine function of the expression product of a gene corresponding to a polynucleotide provided herein. These methods and compositions are particularly useful where the provided novel polynucleotide exhibits no significant or substantial homology to a sequence encoding a gene of known function. Antisense molecules and ribozymes can be constructed from synthetic polynucleotides. Typically, the phosphoramidite method of oligonucleotide synthesis is used. See Beaucage et al., Tet. Lett. (1981) 22:1859 and USPN 4,668,777. Automated devices for synthesis are available to create oligonucleotides using this chemistry. Examples of such devices include Biosearch 8600, Models 392 and 394 by Applied Biosystems, a division of Perkin-Elmer Corp., Foster City, California, USA; and Expedite by Perceptive Biosystems, Framingham, Massachusetts, USA. Synthetic RNA, phosphate analog oligonucleotides, and chemically derivatized oligonucleotides can also be produced, and can be covalently attached to other molecules. RNA oligonucleotides can be synthesized, for example, using RNA phosphoramidites. This method can be performed on an automated synthesizer, such as Applied Biosystems, Models 392 and 394, Foster City, California, USA.

Phosphorothioate oligonucleotides can also be synthesized for antisense construction. A sulfurizing reagent, such as tetraethylthiruam disulfide (TETD) in acetonitrile can be used to convert the internucleotide cyanoethyl phosphite to the phosphorothioate triester within 15 minutes at room temperature. TETD replaces the iodine reagent, while all other reagents used for standard phosphoramidite chemistry remain the same. Such a synthesis method can be automated using Models 392 and 394 by Applied Biosystems, for example.

Oligonucleotides of up to 200 nt can be synthesized, more typically, 100 nt; more typically 50 nt; even more typically, 30 to 40 nt. These synthetic fragments can be annealed and ligated together to construct larger fragments. See, for example, Sambrook et al., supra. Trans-cleaving catalytic RNAs (ribozymes) are RNA molecules possessing endoribonuclease activity. Ribozymes are specifically designed for a particular target, and the target message must contain a specific nucleotide sequence. They are engineered to cleave any RNA species site-specifically in the background of cellular RNA. The cleavage event renders the mRNA unstable and prevents protein expression. Importantly, ribozymes can be used to inhibit expression of a gene of unknown function for the purpose of determining its function in an in vitro or in vivo context, by detecting the phenotypic effect. One commonly used ribozyme motif is the hammerhead, for which the substrate sequence requirements are minimal. Design of the hammerhead ribozyme, as well as therapeutic uses of ribozymes, are disclosed in Usman et al., Current Opin. Struct. Biol. (1996) 6:527. Methods for production of ribozymes, including hairpin structure ribozyme fragments, methods of increasing ribozyme specificity, and the like are known in the art.

The hybridizing region of the ribozyme can be modified or can be prepared as a branched structure as described in Horn and Urdea, Nucleic Acids Res. (1989) 17:6959. The basic structure of the ribozymes can also be chemically altered in ways familiar to those skilled in the art, and chemically synthesized ribozymes can be administered as synthetic oligonucleotide derivatives modified by monomeric units. In a therapeutic context, liposome mediated delivery of ribozymes improves cellular uptake, as described in Birikh et al., Eur. J. Biochem. (1997) 245:1.

Antisense nucleic acids are designed to specifically bind to RNA, resulting in the formation of RNA-DNA or RNA-RNA hybrids, with an arrest of DNA replication, reverse transcription or messenger RNA translation. Antisense polynucleotides based on a selected polynucleotide sequence can interfere with expression of the corresponding gene. Antisense polynucleotides are typically generated within the cell by expression from antisense constructs that contain the antisense strand as the transcribed strand. Antisense polynucleotides based on the disclosed polynucleotides will bind and/or interfere with the translation of mRNA comprising a sequence complementary to the antisense polynucleotide. The expression products of control cells and cells treated with the antisense construct are compared to detect the protein product of the gene corresponding to the polynucleotide upon which the antisense construct is based. The protein is isolated and identified using routine biochemical methods.

Given the extensive background literature and clinical experience in antisense therapy, one skilled in the art can use selected polynucleotides of the invention as additional potential therapeutics. The choice of polynucleotide can be narrowed by first testing them for binding to "hot spot" regions of the genome of cancerous cells. If a polynucleotide is identified as binding to a "hot spot," testing the polynucleotide as an antisense compound in the corresponding cancer cells is warranted.

As an alternative method for identifying function of the gene corresponding to a polynucleotide disclosed herein, dominant negative mutations are readily generated for corresponding proteins that are active as homomultimers. A mutant polypeptide will interact with wild-type polypeptides (made from the other allele) and form a non-functional multimer. Thus, a mutation is in a substrate-binding domain, a catalytic domain, or a cellular localization domain. Preferably, the mutant polypeptide will be overproduced. Point mutations are made that have such an effect. In addition, fusion of different polypeptides of various lengths to the terminus of a protein can yield dominant negative mutants. General strategies are available for making dominant negative mutants (see, e.g., Herskowitz, Nature (1987) 329:219). Such techniques can be used to create loss of function mutations, which are useful for determining protein function.

Polypeptides and Variants Thereof

The polypeptides of the invention include those encoded by the disclosed polynucleotides, as well as nucleic acids that, by virtue of the degeneracy of the genetic code, are not identical in sequence to the disclosed polynucleotides. Thus, the invention includes within its scope a polypeptide encoded by a polynucleotide having the sequence of any one of SEQ ID NOS:1-1485 or a variant thereof. Also included in the invention are the polypeptides comprising the amino acid sequences of SEQ ID NOS:1486-1542.

In general, the term "polypeptide" as used herein refers to both the full length polypeptide encoded by the recited polynucleotide, the polypeptide encoded by the gene represented by the recited polynucleotide, as well as portions or fragments thereof. "Polypeptides" also includes variants of the naturally occurring proteins, where such variants are homologous or substantially similar to the naturally occurring protein, and can be of an origin of the same or different species as the naturally occurring protein (e.g., human, murine, or some other species that naturally expresses the recited polypeptide, usually a mammalian species). In general, variant polypeptides have a sequence that has at least about 80%, usually at least about 90%, and more usually at least about 98% sequence identity with a differentially expressed polypeptide of the invention, as measured by BLAST 2.0 or TeraBLAST using the parameters described above. The variant polypeptides can be naturally or non-naturally glycosylated, i.e., the polypeptide has a glycosylation pattern that differs from the glycosylation pattern found in the corresponding naturally occurring protein.

The invention also encompasses homologs of the disclosed polypeptides (or fragments thereof) where the homologs are isolated from other species, i.e. other animal or plant species, where such homologs, usually mammalian species, e.g. rodents, such as mice, rats; domestic animals, e.g., horse, cow, dog, cat; and humans. By "homolog" is meant a polypeptide having at least about 35%, usually at least about 40% and more usually at least about 60% amino acid sequence identity to a particular differentially expressed protein as identified above, where sequence identity is determined using the BLAST 2.0 or TeraBLAST algorithm, with the parameters described supra.

In general, the polypeptides of the subject invention are provided in a non-naturally occurring environment, e.g. are separated from their naturally occurring environment. In certain embodiments, the subject protein is present in a composition that is enriched for the protein as compared to a control. As such, purified polypeptide is provided, where by purified is meant that the protein is present in a composition that is substantially free of non-differentially expressed polypeptides, where by substantially free is meant that less than 90%, usually less than 60% and more usually less than 50% of the composition is made up of non-differentially expressed polypeptides.

Also within the scope of the invention are variants; variants of polypeptides include mutants, fragments, and fusions. Mutants can include amino acid substitutions, additions or deletions. The amino acid substitutions can be conservative amino acid substitutions or substitutions to eliminate non-essential amino acids, such as to alter a glycosylation site, a phosphorylation site or an acetylation site, or to minimize misfolding by substitution or deletion of one or more cysteine residues that are not necessary for function. Conservative amino acid substitutions are those that preserve the general charge, hydrophobicity/ hydrophilicity, and/or steric bulk of the amino acid substituted. Variants can be designed so as to retain or have enhanced biological activity of a particular region of the protein (e.g., a functional domain and/or, where the polypeptide is a member of a protein family, a region associated with a consensus sequence). Selection of amino acid alterations for production of variants can be based upon the accessibility (interior vs. exterior) of the amino acid (see, e.g., Go et al, Int. J. Peptide Protein Res. (1980) 15:211), the thermostability of the variant polypeptide (see, e.g., Querol et al., Prot. Eng. (1996) 9:265), desired glycosylation sites (see, e.g., Olsen and Thomsen, J. Gen. Microbiol. (1991) 137:579), desired disulfide bridges (see, e.g., Clarke et al., Biochemistry (1993) 32:4322; and Wakarchuk et al., Protein Eng. (1994) 7:1379), desired metal binding sites (see, e.g., Toma et al., Biochemistry (1991) 30:97, and Haezerbrouck et al., Protein Eng. (1993) 6:643), and desired substitutions within proline loops (see, e.g., Masul et al., Appl. Env. Microbiol. (1994) 60:3579). Cysteine-depleted muteins can be produced as disclosed in USPN 4,959,314.

Variants also include fragments of the polypeptides disclosed herein, particularly haptens, biologically active fragments, and/or fragments corresponding to functional domains. Fragments of interest will typically be at least about 10 aa to at least about 15 aa in length, usually at least about 50 aa in length, and can be as long as 300 aa in length or longer, but will usually not exceed about 1000 aa in length, where the fragment will have a stretch of amino acids that is identical to a polypeptide encoded by a polynucleotide having a sequence of any SEQ ID NOS:1-1485, a polypeptide comprising a sequence of at least one of SEQ ID NOS:1486-1542, or a homolog thereof. The protein variants described herein are encoded by polynucleotides that are within the scope of the invention. The genetic code can be used to select the appropriate codons to construct the corresponding variants.

A fragment of a subject polypeptide is, for example, a polypeptide having an amino acid sequence which is a portion of a subject polypeptide e.g. a polypeptide encoded by a subject polynucleotide that is identified by any one of the sequence the sequence listing or its complement. The polypeptide fragments of the invention are preferably at least about 9 aa, at least about 15 aa, and more preferably at least about 20 aa, still more preferably at least about 30 aa, and even more preferably, at least about 40 aa, at least about 50 aa, at least about 75 aa, at least about 100 aa, at least about 125 aa or at least about 150 aa in length. A fragment "at least 20 aa in length," for example, is intended to include 20 or more contiguous amino acids from, for example, the polypeptide encoded by a cDNA, in a cDNA clone contained in a deposited library, or a nucleotide sequence shown in the sequence listing or the complementary stand thereof. In this context "about" includes the particularly recited value or a value larger or smaller by several (5, 4, 3, 2, or 1) amino acids. These polypeptide fragments have uses that include, but are not limited to, production of antibodies as discussed herein. Of course, larger fragments (e.g., at least 150, 175, 200, 250, 500, 600, 1000, or 2000 amino acids in length) are also encompassed by the invention.

Moreover, representative examples of polypeptides fragments of the invention (useful in, for example, as antigens for antibody production), include, for example, fragments comprising, or alternatively consisting of, a sequence from about amino acid number 1-10, 5-10, 10-20, 21-31, 31-40, 41-61, 61-81, 91-120, 121-140, 141-162, 162-200, 201-240, 241-280, 281- 320, 321-360, 360-400, 400-450, 451-500, 500-600, 600-700, 700-800, 800-900 and the like. In this context "about" includes the particularly recited range or a range larger or smaller by several (5, 4, 3, 2, or 1) amino acids, at either terminus or at both termini. In some embodiments, these fragments has a functional activity (e.g., biological activity) whereas in other embodiments, these fragments may be used to make an antibody.

Further polypeptide variants may are described in PCT publications WO/00-55173, WO/01-07611 and WO/02-16429

Computer-Related Embodiments

In general, a library of polynucleotides is a collection of sequence information, which information is provided in either biochemical form (e.g., as a collection of polynucleotide molecules), or in electronic form (e.g., as a collection of polynucleotide sequences stored in a computer-readable form, as in a computer system and/or as part of a computer program). The sequence information of the polynucleotides can be used in a variety of ways, e.g., as a resource for gene discovery, as a representation of sequences expressed in a selected cell type (e.g., cell type markers), and/or as markers of a given disease or disease state. In general, a disease marker is a representation of a gene product that is present in all cells affected by disease either at an increased or decreased level relative to a normal cell (e.g., a cell of the same or similar type that is not substantially affected by disease). For example, a polynucleotide sequence in a library can be a polynucleotide that represents an mRNA, polypeptide, or other gene product encoded by the polynucleotide, that is either overexpressed or underexpressed in a breast ductal cell affected by cancer relative to a normal (i.e., substantially disease-free) breast cell.

The nucleotide sequence information of the library can be embodied in any suitable form, e.g., electronic or biochemical forms. For example, a library of sequence information embodied in electronic form comprises an accessible computer data file (or, in biochemical form, a collection of nucleic acid molecules) that contains the representative nucleotide sequences of genes that are differentially expressed (e.g., overexpressed or underexpressed) as between, for example, i) a cancerous cell and a normal cell; ii) a cancerous cell and a dysplastic cell; iii) a cancerous cell and a cell affected by a disease or condition other than cancer; iv) a metastatic cancerous cell and a normal cell and/or non-metastatic cancerous cell; v) a malignant cancerous cell and a non-malignant cancerous cell (or a normal cell) and/or vi) a dysplastic cell relative to a normal cell. Other combinations and comparisons of cells affected by various diseases or stages of disease will be readily apparent to the ordinarily skilled artisan. Biochemical embodiments of the library include a collection of nucleic acids that have the sequences of the genes in the library, where the nucleic acids can correspond to the entire gene in the library or to a fragment thereof, as described in greater detail below.

The polynucleotide libraries of the subject invention generally comprise sequence information of a plurality of polynucleotide sequences, where at least one of the polynucleotides has a sequence of any of SEQ ID NOS:1-1485. By plurality is meant at least 2, usually at least 3 and can include up to all of SEQ ID NOS:1-1485. The length and number of polynucleotides in the library will vary with the nature of the library, e.g., if the library is an oligonucleotide array, a cDNA array, a computer database of the sequence information, etc.

Where the library is an electronic library, the nucleic acid sequence information can be present in a variety of media. "Media" refers to a manufacture, other than an isolated nucleic acid molecule, that contains the sequence information of the present invention. Such a manufacture provides the genome sequence or a subset thereof in a form that can be examined by means not directly applicable to the sequence as it exists in a nucleic acid. For example, the nucleotide sequence of the present invention, e.g. the nucleic acid sequences of any of the polynucleotides of SEQ ID NOS:1-1485, can be recorded on computer readable media, e.g. any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as a floppy disc, a hard disc storage medium, and a magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media. One of skill in the art can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising a recording of the present sequence information. "Recorded" refers to a process for storing information on computer readable medium, using any such methods as known in the art. Any convenient data storage structure can be chosen, based on the means used to access the stored information. A variety of data processor programs and formats can be used for storage, e.g. word processing text file, database format, etc. In addition to the sequence information, electronic versions of the libraries of the invention can be provided in conjunction or connection with other computer-readable information and/or other types of computer-readable files (e.g., searchable files, executable files, etc, including, but not limited to, for example, search program software, etc.).

By providing the nucleotide sequence in computer readable form, the information can be accessed for a variety of purposes. Computer software to access sequence information is publicly available. For example, the gapped BLAST (Altschul et al. Nucleic Acids Res. (1997) 25:3389-3402) and BLAZE (Brutlag et al. Comp. Chem. (1993) 17:203) search algorithms on a Sybase system, or the TeraBLAST (TimeLogic, Crystal Bay, Nevada) program optionally running on a specialized computer platform available from TimeLogic, can be used to identify open reading frames (ORFs) within the genome that contain homology to ORFs from other organisms.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based system are suitable for use in the present invention. The data storage means can comprise any manufacture comprising a recording of the present sequence information as described above, or a memory access means that can access such a manufacture.

"Search means" refers to one or more programs implemented on the computer-based system, to compare a target sequence or target structural motif, or expression levels of a polynucleotide in a sample, with the stored sequence information. Search means can be used to identify fragments or regions of the genome that match a particular target sequence or target motif. A variety of known algorithms are publicly known and commercially available, e.g. MacPattern (EMBL), BLASTN and BLASTX (NCBI), TeraBLAST (TimeLogic, Crystal Bay, Nevada). A "target sequence" can be any polynucleotide or amino acid sequence of six or more contiguous nucleotides or two or more amino acids, preferably from about 10 to 100 amino acids or from about 30 to 300 nt A variety of comparing means can be used to accomplish comparison of sequence information from a sample (e.g., to analyze target sequences, target motifs, or relative expression levels) with the data storage means. A skilled artisan can readily recognize that any one of the publicly available homology search programs can be used as the search means for the computer based systems of the present invention to accomplish comparison of target sequences and motifs. Computer programs to analyze expression levels in a sample and in controls are also known in the art.

A "target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration that is formed upon the folding of the target motif, or on consensus sequences of regulatory or active sites. There are a variety of target motifs known in the art. Protein target motifs include, but arc not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, hairpin structures, promoter sequences and other expression elements such as binding sites for transcription factors.

A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. One format for an output means ranks the relative expression levels of different polynucleotides. Such presentation provides a skilled artisan with a ranking of relative expression levels to determine a gene expression profile.

As discussed above, the "library" of the invention also encompasses biochemical libraries of the polynucleotides of SEQ ID NOS:1-1485 , e.g., collections of nucleic acids representing the provided polynucleotides. The biochemical libraries can take a variety of forms, e.g., a solution of cDNAs, a pattern of probe nucleic acids stably associated with a surface of a solid support (i.e., an array) and the like. Of particular interest are nucleic acid arrays in which one or more of SEQ ID NOS:1-1485 is represented on the array. By array is meant a an article of manufacture that has at least a substrate with at least two distinct nucleic acid targets on one of its surfaces, where the number of distinct nucleic acids can be considerably higher, typically being at least 10, usually at least 20, and often at least 25 distinct nucleic acid molecules. A variety of different array formats have been developed and are known to those of skill in the art. The arrays of the subject invention find use in a variety of applications, including gene expression analysis, drug screening, mutation analysis and the like, as disclosed in the above-listed exemplary patent documents.

In addition to the above nucleic acid libraries, analogous libraries of polypeptides are also provided, where the polypeptides of the library will represent at least a portion of the polypeptides encoded by a gene corresponding to one or more of SEQ ID NOS:1-1485.

Utilities

The polynucleotides of the invention are useful in a variety of applications. Exemplary utilies of the polynucleotides of the invention are described below.

Construction of Larger Molecules: Recombinant DNAs and Nucleic Acid Multimers. In one embodiment of particular interest, the polynucleotides described herein as useful as the building blocks for larger molecules. In one example, the polynucleotide is a component of a larger cDNA molecule which in turn can be adapted for expression in a host cell (*e.g.*, a bacterial or eukaryotic (*e.g.*, yeast or mammalian) host cell). The cDNA can include, in addition to the polypeptide encoded by the starting material polynucleotide (*i.e.*, a polynucleotide described herein), an amino acid sequence that is heterologous to the polypeptide encoded by the polynucleotide described herein (*e.g.*, as in a sequence encoding a fusion protein). In some embodiments, the polynucleotides described herein is used as starting material polynucleotide for synthesizing all or a portion of the gene to which the described polynucleotide corresponds. For example, a DNA molecule encoding a full-length human polypeptide can be constructed using a polynucleotide described herein as starting material.

In another embodiment, the polynucleotides of the invention are used in nucleic acid multimers. Nucleic acid multimers can be linear or branched polymers of the same repeating single-stranded oligonucleotide unit or different single-stranded oligonucleotide units. Where the molecules are branched, the multimers are generally described as either "fork" or "comb" structures. The oligonucleotide units of the multimer may be composed of RNA, DNA, modified nucleotides or combinations thereof. At least one of the units has a sequence, length, and composition that permits it to bind specifically to a first single-stranded nucleotide sequence of interest, typically analyte or an oligonucleotide bound to the analyte. In order to achieve such specificity and stability, this unit will normally be 15 to 50 nt, preferably 15 to 30 nt, in length and have a GC content in the range of 40% to 60%. In addition to such unit(s), the multimer includes a multiplicity of units that are capable of hybridizing specifically and stably to a second single-stranded nucleotide of interest, typically a labeled oligonucleotide or another multimer. These units will also normally be 15 to 50 nt, preferably 15 to 30 nt, in length and have a GC content in the range of 40% to 60%. When a multimer is designed to be hybridized to another multimer, the first and second oligonucleotide units are heterogeneous (different). One or more of the polynucleotides described herein, or a portion of a polynucleotide described herein, can be used as a repeating unit of such nucleic acid multimers.

The total number of oligonucleotide units in the multimer will usually be in the range of 3 to 50, more usually 10 to 20. In multimers in which the unit that hybridizes to the nucleotide sequence of interest is different from the unit that hybridizes to the labeled oligonucleotide, the number ratio of the latter to the former will usually be 2:1 to 30:1, more usually 5:1 to 20:1, and-preferably 10:1 to 15:1.

The oligonucleotide units of the multimer may be covalently linked directly to each other through phosphodiester bonds or through interposed linking agents such as nucleic acid, amino acid, carbohydrate or polyol bridges, or through other cross-linking agents that are capable of cross-linking nucleic acid or modified nucleic acid strands. The site(s) of linkage may be at the ends of the unit (in either normal 3,-5' orientation or randomly oriented) and/or at one or more internal nucleotides in the strand. In linear multimers the individual units are linked end-to-end to form a linear polymer. In one type of branched multimer three or more oligonucleotide units emanate from a point of origin to form a branched structure. The point of origin may be another oligonucleotide unit or a multifunctional molecule to which at least three units can be covalently bound. In another type, there is an oligonucleotide unit backbone with one or more pendant oligonucleotide units. These latter-type multimers are "fork-like", "comb-like" or combination "fork-" and "comb-like" in structure. The pendant units will normally depend from a modified nucleotide or other organic moiety having appropriate functional groups to which oligonucleotides may be conjugated or otherwise attached. The multimer may be totally linear, totally branched, or a combination of linear and branched portions. Preferably there will be at least two branch points in the multimer, more preferably at least 3, preferably 5 to 10. The multimer may include one or more segments of double-stranded sequences.

Multimeric nucleic acid molecules are useful in amplifying the signal that results from hybridization of one the first sequence of the multimeric molecule to a target sequence. The amplification is theoretically proportional to the number of iterations of the second segment.

Without being held to theory, forked structures of greater than about eight branches exhibited steric hindrance which inhibited binding of labeled probes to the multimer. On the other hand, comb structures exhibit little or no steric problems and are thus a preferred type of branched multimer. For a description of branched nucleic acid multimers of both the fork and comb types, as well as methods of use and synthesis, see, *e.g.*, U.S. Pat. Nos. 5,124,246 (fork-type structures); 5,710,264 (synthesis of comb structures); and 5,849,481.

Use of Polynucleotide Probes in Mapping, and in Tissue Profiling. Polynucleotide probes, generally comprising at least 12 contiguous nt of a polynucleotide as shown in the Sequence Listing, are used for a variety of purposes, such as chromosome mapping of the polynucleotide and detection of transcription levels. Additional disclosure about preferred regions of the disclosed polynucleotide sequences is found in the Examples. A probe that hybridizes specifically to a polynucleotide disclosed herein should provide a detection signal at least 5-, 10-, or 20-fold higher than the background hybridization provided with other unrelated sequences.

Detection of Expression Levels. Nucleotide probes are used to detect expression of a gene corresponding to the provided polynucleotide. In Northern blots, mRNA is separated electrophoretically and contacted with a probe. A probe is detected as hybridizing to an mRNA species of a particular size. The amount of hybridization is quantitated to determine relative amounts of expression, for example under a particular condition. Probes are used for in situ hybridization to cells to detect expression. Probes can also be used in vivo for diagnostic detection of hybridizing sequences. Probes are typically labeled with a radioactive isotope. Other types of detectable labels can be used such as chromophores, fluors, and enzymes. Other examples of nucleotide hybridization assays are described in WO92/02526 and USPN 5,124,246.

Alternatively, the Polymerase Chain Reaction (PCR) is another means for detecting small amounts of target nucleic acids (see, e.g., Mullis et al., Meth. Enzymol. (1987) 155:335; USPN 4,683,195; and USPN 4,683,202). Two primer polynucleotides nucleotides that hybridize with the target nucleic acids are used to prime the reaction. The primers can be composed of sequence within or 3' and 5' to the polynucleotides of the Sequence Listing. Alternatively, if the primers are 3' and 5' to these polynucleotides, they need not hybridize to them or the complements. After amplification of the target with a thermostable polymerase, the amplified target nucleic acids can be detected by methods known in the art, e.g., Southern blot. mRNA or cDNA can also be detected by traditional blotting techniques (e.g., Southern blot, Northern blot, etc.) described in Sambrook et al., "Molecular Cloning: A Laboratory Manual" (New York, Cold Spring Harbor Laboratory, 1989) (e.g., without PCR amplification). In general, mRNA or cDNA generated from mRNA using a polymerase enzyme can be purified and separated using gel electrophoresis, and transferred to a solid support, such as nitrocellulose. The solid support is exposed to a labeled probe, washed to remove any unhybridized probe, and duplexes containing the labeled probe are detected.

Mapping. Polynucleotides of the present invention can be used to identify a chromosome on which the corresponding gene resides. Such mapping can be useful in identifying the function of the polynucleotide-related gene by its proximity to other genes with known function. Function can also be assigned to the polynucleotide-related gene when particular syndromes or diseases map to the same chromosome. For example, use of polynucleotide probes in identification and quantification of nucleic acid sequence aberrations is described in USPN 5,783,387. An exemplary mapping method is fluorescence in situ hybridization (FISH), which facilitates comparative genomic hybridization to allow total genome assessment of changes in relative copy number of DNA sequences (see, e.g., Valdes et al., Methods in Molecular Biology (1997) 68:1). Polynucleotides can also be mapped to particular chromosomes using, for example, radiation hybrids or chromosome-specific hybrid panels. See Leach et al., Advances in Genetics, (1995) 33:63-99; Walter et al., Nature Genetics (1994) 7:22; Walter and Goodfellow, Trends in Genetics (1992) 9:352. Panels for radiation hybrid mapping are available from Research Genetics, Inc., Huntsville, Alabama, USA. Databases for markers using various panels are available via the world wide web at sites supported by the Stanford Human Genome Center (Stanford University) and the Whitehead Institute for Biomedical Research/MIT Center for Genome Research. The statistical program RHMAP can be used to construct a map based on the data from radiation hybridization with a measure of the relative likelihood of one order versus another. RHMAP is available via the world wide web at a site supported by the University of Michigan. In addition, commercial programs are available for identifying regions of chromosomes commonly associated with disease, such as cancer.

Tissue Typing or Profiling. Expression of specific mRNA corresponding to the provided polynucleotides can vary in different cell types and can be tissue-specific. This variation of mRNA levels in different cell types can be exploited with nucleic acid probe assays to determine tissue types. For example, PCR, branched DNA probe assays, or blotting techniques utilizing nucleic acid probes substantially identical or complementary to polynucleotides listed in the Sequence Listing can determine the presence or absence of the corresponding cDNA or mRNA.

Tissue typing can be used to identify the developmental organ or tissue source of a metastatic lesion by identifying the expression of a particular marker of that organ or tissue. If a polynucleotide is expressed only in a specific tissue type, and a metastatic lesion is found to express that polynucleotide, then the developmental source of the lesion has been identified. Expression of a particular polynucleotide can be assayed by detection of either the corresponding mRNA or the protein product. As would be readily apparent to any forensic scientist, the sequences disclosed herein are useful in differentiating human tissue from non-human tissue. In particular, these sequences are useful to differentiate human tissue from bird, reptile, and amphibian tissue, for example.

Use of Polymorphisms. A polynucleotide of the invention can be used in forensics, genetic analysis, mapping, and diagnostic applications where the corresponding region of a gene is polymorphic in the human population. Any means for detecting a polymorphism in a gene can be used, including, but not limited to electrophoresis of protein polymorphic variants, differential sensitivity to restriction enzyme cleavage, and hybridization to allele-specific probes.

Antibody Production. The present invention further provides antibodies, which may be isolated antibodies, that are specific for a polypeptide encoded by a polynucleotide described herein (*e.g.*, a polypeptide encoded by a sequence corresponding to SEQ ID NOS:1-1485, a polypeptide comprising an amino acid sequence of SEQ ID NOS:1486-1542). Antibodies can be provided in a composition comprising the antibody and a buffer and/or a pharmaceutically acceptable excipient. Antibodies specific for a polypeptide associated with prostate cancer are useful in a variety of diagnostic and therapeutic methods, as discussed in detail herein.

Expression products of a polynucleotide of the invention, as well as the corresponding mRNA, cDNA, or complete gene, can be prepared and used for raising antibodies for experimental, diagnostic, and therapeutic purposes. For polynucleotides to which a corresponding gene has not been assigned, this provides an additional method of identifying the corresponding gene. The polynucleotide or related cDNA is expressed as described above, and antibodies are prepared. These antibodies are specific to an epitope on the polypeptide encoded by the polynucleotide, and can precipitate or bind to the corresponding native protein in a cell or tissue preparation or in a cell-free extract of an in vitro expression system.

Methods for production of antibodies that specifically bind a selected antigen are well known in the art. Immunogens for raising antibodies can be prepared by mixing a polypeptide encoded by a polynucleotide of the invention with an adjuvant, and/or by making fusion proteins with larger immunogenic proteins. Polypeptides can also be covalently linked to other larger immunogenic proteins, such as keyhole limpet hemocyanin. Immunogens are typically administered intradermally, subcutaneously, or intramuscularly to experimental animals such as rabbits, sheep, and mice, to generate antibodies. Monoclonal antibodies can be generated by isolating spleen cells and fusing myeloma cells to form hybridomas. Alternatively, the selected polynucleotide is administered directly, such as by intramuscular injection, and expressed in vivo. The expressed protein generates a variety of protein-specific immune responses, including production of antibodies, comparable to administration of the protein.

Preparations of polyclonal and monoclonal antibodies specific for polypeptides encoded by a selected polynucleotide are made using standard methods known in the art. The antibodies specifically bind to epitopes present in the polypeptides encoded by polynucleotides disclosed in the Sequence Listing. Typically, at least 6, 8, 10, or 12 contiguous amino acids are required to form an epitope. Epitopes that involve non-contiguous amino acids may require a longer polypeptide, e.g., at least 15, 25, or 50 amino acids. Antibodies that specifically bind to human polypeptides encoded by the provided polypeptides should provide a detection signal at least 5-, 10-, or 20-fold higher than a detection signal provided with other proteins when used in Western blots or other immunochemical assays. Preferably, antibodies that specifically bind polypeptides contemplated by the invention do not bind to other proteins in immunochemical assays at detectable levels and can immunoprecipitate the specific polypeptide from solution.

The invention also contemplates naturally occurring antibodies specific for a polypeptide of the invention. For example, serum antibodies to a polypeptide of the invention in a human population can be purified by methods well known in the art, e.g., by passing antiserum over a column to which the corresponding selected polypeptide or fusion protein is bound. The bound antibodies can then be eluted from the column, for example, using a buffer with a high salt concentration.

In addition to the antibodies discussed above, the invention also contemplates genetically engineered antibodies antibodies (e.g., chimeric antibodies, humanized antibodies, human antibodies produced by a transgenic animal (e.g., a transgenic mouse such as the XenomousTM), antibody derivatives (e.g., single chain antibodies, antibody fragments (e.g., Fab, etc.)), according to methods well known in the art.

The invention also contemplates other molecules that can specifically bind a polynucleotide or polypeptide of the invention. Examples of such molecules include, but are not necessarily limited to, single-chain binding proteins (*e.g.*, mono- and multi-valent single chain antigen binding proteins (*see, e.g.*, U.S. Patent Nos. 4,704,692; 4,946,778; 4,946,778; 6,027,725; 6,121,424)), oligonucleotide-based synthetic antibodies (*e.g.*, oligobodies (*see, e.g.*, Radrizzani *et al.*, *Medicina* (B Aires) (1999) 59:753-8; Radrizzani *et al.*, *Medicina* (B Aires) (2000) 60(Suppl 2):55-60)), aptamers (*see, e.g.*, Gening *et al.*, *Biotechniques* (2001) 3:828, 830, 832, 834; Cox and Ellington, *Bioorg. Med. Chem.* (2001) 9:2525-31), and the like.

Polynucleotides or Arrays for Diagnostics.

Polynucleotide arrays provide a high throughput technique that can assay a large number of polynucleotides in a sample. This technology can be used as a diagnostic and as tool to test for differential expression expression, e.g., to determine function of an encoded protein. A variety of methods of producing arrays, as well as variations of these methods, are known in the art and contemplated for use in the invention. For example, arrays can be created by spotting polynucleotide probes onto a substrate (e.g., glass, nitrocellulose, etc.) in a two-dimensional matrix or array having bound probes. The probes can be bound to the substrate by either covalent bonds or by non-specific interactions, such as hydrophobic interactions. Samples of polynucleotides can be detectably labeled (e.g., using radioactive or fluorescent labels) and then hybridized to the probes. Double stranded polynucleotides, comprising the labeled sample polynucleotides bound to probe polynucleotides, can be detected once the unbound portion of the sample is washed away. Alternatively, the polynucleotides of the test sample can be immobilized on the array, and the probes detectably labeled. Techniques for constructing arrays and methods of using these arrays are described in, for example, Schena et al. (1996) Proc Natl Acad Sci U S A. 93(20):10614-9; Schena et al. (1995) Science 270(5235):467-70; Shalon et al. (1996) Genome Res. 6(7):639-45, USPN 5,807,522, EP 799 897; WO 97/29212; WO 97/27317; EP 785 280; WO 97/02357; USPN 5,593,839; USPN 5,578,832; EP 728 520; USPN 5,599,695; EP 721 016; USPN 5,556,752; WO 95/22058; and USPN 5,631,734.

Arrays can be used to, for example, examine differential expression of genes and can be used to determine gene function. For example, arrays can be used to detect differential expression of a gene corresponding to a polynucleotide of the invention, where expression is compared between a test cell and control cell (e.g., cancer cells and normal cells). For example, high expression of a particular message in a cancer cell, which is not observed in a corresponding normal cell, can indicate a cancer specific gene product. Exemplary uses of arrays are further described in, for example, Pappalarado et al., Sem. Radiation Oncol. (1998) 8:217; and Ramsay Nature Biotechnol. (1998) 16:40. Furthermore, many variations on methods of detection using arrays are well within the skill in the art and within the scope of the present invention. For example, rather than immobilizing the probe to a solid support, the test sample can be immobilized on a solid support which is then contacted with the probe.

Differential Expression in Diagnosis

The polynucleotides of the invention can also be used to detect differences in expression levels between two cells, e.g., as a method to identify abnormal or diseased tissue in a human. For polynucleotides corresponding to profiles of protein families, the choice of tissue can be selected according to the putative biological function. In general, the expression of a gene corresponding to a specific polynucleotide is compared between a first tissue that is suspected of being diseased and a second, normal tissue of the human. The tissue suspected of being abnormal or diseased can be derived from a different tissue type of the human, but preferably it is derived from the same tissue type; for example, an intestinal polyp or other abnormal growth should be compared with normal intestinal tissue. The normal tissue can be the same tissue as that of the test sample, or any normal tissue of the patient, especially those that express the polynucleotide-related gene of interest (e.g., brain, thymus, testis, heart, prostate, placenta, spleen, small intestine, skeletal muscle, pancreas, and the mucosal lining of the colon). A difference between the polynucleotide-related gene, mRNA, or protein in the two tissues which are compared, for example, in molecular weight, amino acid or nucleotide sequence, or relative abundance, indicates a change in the gene, or a gene which regulates it, in the tissue of the human that was suspected of being diseased. Examples of detection of differential expression and its use in diagnosis of cancer are described in USPNs 5,688,641 and 5,677,125.

A genetic predisposition to disease in a human can also be detected by comparing expression levels of an mRNA or protein corresponding to a polynucleotide of the invention in a fetal tissue with levels associated in normal fetal tissue. Fetal tissues that are used for this purpose include, but are not limited to, amniotic fluid, chorionic villi, blood, and the blastomere of an in vitro-fertilized embryo. The comparable normal polynucleotide-related gene is obtained from any tissue. The mRNA or protein is obtained from a normal tissue of a human in which the polynucleotide-related gene is expressed. Differences such as alterations in the nucleotide sequence or size of the same product of the fetal polynucleotide-related gene or mRNA, or alterations in the molecular weight, amino acid sequence, or relative abundance of fetal protein, can indicate a germline mutation in the polynucleotide-related gene of the fetus, which indicates a genetic predisposition to disease. In general, diagnostic, prognostic, and other methods of the invention based on differential expression involve detection of a level or amount of a gene product, particularly a differentially expressed gene product, in a test sample obtained from a patient suspected of having or being susceptible to a disease (e.g., breast cancer, prostate cancer, lung cancer, colon cancer and/or metastatic forms thereof), and comparing the detected levels to those levels found in normal cells (e.g., cells substantially unaffected by cancer) and/or other control cells (e.g., to differentiate a cancerous cell from a cell affected by dysplasia). Furthermore, the severity of the disease can be assessed by comparing the detected levels of a differentially expressed gene product with those levels detected in samples representing the levels of differentially expressed gene product associated with varying degrees of severity of disease. It should be noted that use of the term "diagnostic" herein is not necessarily meant to exclude "prognostic" or "prognosis," but rather is used as a matter of convenience.

The term "differentially expressed gene" is generally intended to encompass a polynucleotide that can, for example, include an open reading frame encoding a gene product (e.g., a polypeptide), and/or introns of such genes and adjacent 5' and 3' non-coding nucleotide sequences involved in the regulation of expression, up to about 20 kb beyond the coding region, but possibly further in either direction. The gene can be introduced into an appropriate vector for extrachromosomal maintenance or for integration into a host genome. In general, a difference in expression level associated with a decrease in expression level of at least about 25%, usually at least about 50% to 75%, more usually at least about 90% or more is indicative of a differentially expressed gene of interest, i.e., a gene that is underexpressed or down-regulated in the test sample relative to a control sample. Furthermore, a difference in expression level associated with an increase in expression of at least about 25%, usually at least about 50% to 75%, more usually at least about 90% and can be at least about 1½-fold, usually at least about 2-fold to about 10-fold, and can be about 100-fold to about 1,000-fold increase relative to a control sample is indicative of a differentially expressed gene of interest, i.e., an overexpressed or up-regulated gene.

"Differentially expressed polynucleotide" as used herein means a nucleic acid molecule (RNA or DNA) comprising a sequence that represents a differentially expressed gene, e.g., the differentially expressed polynucleotide comprises a sequence (e.g., an open reading frame encoding a gene product) that uniquely identifies a differentially expressed gene so that detection of the differentially expressed polynucleotide in a sample is correlated with the presence of a differentially expressed gene in a sample. "Differentially expressed polynucleotide" is also meant to encompass fragments of the disclosed polynucleotides, e.g., fragments retaining biological activity, as well as nucleic acids homologous, substantially similar, or substantially identical (e.g., having about 90% sequence identity) to the disclosed polynucleotides.

Methods of the subject invention useful in diagnosis or prognosis typically involve comparison of the abundance of a selected differentially expressed gene product in a sample of interest with that of a control to determine any relative differences in the expression of the gene product, where the difference can be measured qualitatively and/or quantitatively. Quantitation can be accomplished, for example, by comparing the level of expression product detected in the sample with the amounts of product present in a standard curve. A comparison can be made visually; by using a technique such as densitometry, with or without computerized assistance; by preparing a representative library of cDNA clones of mRNA isolated from a test sample, sequencing the clones in the library to determine that number of cDNA clones corresponding to the same gene product, and analyzing the number of clones corresponding to that same gene product relative to the number of clones of the same gene product in a control sample; or by using an array to detect relative levels of hybridization to a selected sequence or set of sequences, and comparing the hybridization pattern to that of a control. The differences in expression are then correlated with the presence or absence of an abnormal expression pattern. A variety of different methods for determining the nucleic acid abundance in a sample are known to those of skill in the art (see, e.g., WO 97/27317).

In general, diagnostic assays of the invention involve detection of a gene product of a polynucleotide sequence (e.g., mRNA or polypeptide) that corresponds to a sequence of SEQ ID NOS:1-1485. The patient from whom the sample is obtained can be apparently healthy, susceptible to disease (e.g., as determined by family history or exposure to certain environmental factors), or can already be identified as having a condition in which altered expression of a gene product of the invention is implicated.

Diagnosis can be determined based on detected gene product expression levels of a gene product encoded by at least one, preferably at least two or more, at least 3 or more, or at least 4 or more of the polynucleotides having a sequence set forth in SEQ ID NOS:1-1485, and can involve detection of expression of genes corresponding to all of SEQ ID NOS:1-1485 and/or additional sequences that can serve as additional diagnostic markers and/or reference sequences. Where the diagnostic method is designed to detect the presence or susceptibility of a patient to cancer, the assay preferably involves detection of a gene product encoded by a gene corresponding to a polynucleotide that is differentially expressed in cancer. Examples of such differentially expressed polynucleotides are described in the Examples below. Given the provided polynucleotides and information regarding their relative expression levels provided herein, assays using such polynucleotides and detection of their expression levels in diagnosis and prognosis will be readily apparent to the ordinarily skilled artisan.

Any of a variety of detectable labels can be used in connection with the various embodiments of the diagnostic methods of the invention. Suitable detectable labels include fluorochromes,(e.g. fluorescein isothiocyanate (FITC), rhodamine, Texas Red, phycoerythrin, allophycocyanin, 6-carboxyfluorescein (6-FAM), 2',7'-dimethoxy-4',5'-dichloro-6-carboxyfluorescein, 6-carboxy-X-rhodamine (ROX), 6-carboxy-2',4',7',4,7-hexachlorofluorescein (HEX), 5-carboxyfluorescein (5-FAM) or N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA)), radioactive labels, (e.g. 32P, 35S, 3H, etc.), and the like. The detectable label can involve a two stage systems (e.g., biotin-avidin, hapten-anti-hapten antibody, etc.).

Reagents specific for the polynucleotides and polypeptides of the invention, such as antibodies and nucleotide probes, can be supplied in a kit for detecting the presence of an expression product in a biological sample. The kit can also contain buffers or labeling components, as well as instructions for using the reagents to detect and quantify expression products in the biological sample. Exemplary embodiments of the diagnostic methods of the invention are described below in more detail.

Polypeptide detection in diagnosis. In one embodiment, the test sample is assayed for the level of a differentially expressed polypeptide, such as a polypeptide of a gene corresponding to SEQ ID NOS:1-1485 and/or a polypeptide comprising a sequence of SEQ ID NO:1486-1542. Diagnosis can be accomplished using any of a number of methods to determine the absence or presence or altered amounts of the differentially expressed polypeptide in the test sample. For example, detection can utilize staining of cells or histological sections with labeled antibodies, performed in accordance with conventional methods. Cells can be permeabilized to stain cytoplasmic molecules. In general, antibodies that specifically bind a differentially expressed polypeptide of the invention are added to a sample, and incubated for a period of time sufficient to allow binding to the epitope, usually at least about 10 minutes. The antibody can be detectably labeled for direct detection (e.g., using radioisotopes, enzymes, fluorescers, chemiluminescers, and the like), or can be used in conjunction with a second stage antibody or reagent to detect binding (e.g., biotin with horseradish peroxidase-conjugated avidin, a secondary antibody conjugated to a fluorescent compound, e.g. fluorescein, rhodamine, Texas red, etc.). The absence or presence of antibody binding can be determined by various methods, including flow cytometry of dissociated cells, microscopy, radiography, scintillation counting, etc. Any suitable alternative methods of qualitative or quantitative detection of levels or amounts of differentially expressed polypeptide can be used, for example, ELISA, western blot, immunoprecipitation, radioimmunoassay, etc.

mRNA detection. The diagnostic methods of the invention can also or alternatively involve detection of mRNA encoded by a gene corresponding to a differentially expressed polynucleotide of the invention. Any suitable qualitative or quantitative methods known in the art for detecting specific mRNAs can be used. mRNA can be detected by, for example, in situ hybridization in tissue sections, by reverse transcriptase-PCR, or in Northern blots containing poly A+ mRNA. One of skill in the art can readily use these methods to determine differences in the size or amount of mRNA transcripts between two samples. mRNA expression levels in a sample can also be determined by generation of a library of expressed sequence tags (ESTs) from the sample, where the EST library is representative of sequences present in the sample (Adams et al., (1991) Science 252:1651). Enumeration of the relative representation of ESTs within the library can be used to approximate the relative representation of the gene transcript within the starting sample. The results of EST analysis of a test sample can then be compared to EST analysis of a reference sample to determine the relative expression levels of a selected polynucleotide, particularly a polynucleotide corresponding to one or more of the differentially expressed genes described herein. Alternatively, gene expression in a test sample can be performed using serial analysis of gene expression (SAGE) methodology (e.g., Velculescu et al., Science (1995) 270:484) or differential display (DD) methodology (see, e.g., USPN 5,776,683 and USPN 5,807,680).

Alternatively, gene expression can be analyzed using hybridization analysis. Oligonucleotides or cDNA can be used to selectively identify or capture DNA or RNA of specific sequence composition, and the amount of RNA or cDNA hybridized to a known capture sequence determined qualitatively or quantitatively, to provide information about the relative representation of a particular message within the pool of cellular messages in a sample. Hybridization analysis can be designed to allow for concurrent screening of the relative expression of hundreds to thousands of genes by using, for example, array-based technologies having high density formats, including filters, microscope slides, or microchips, or solution-based technologies that use spectroscopic analysis (e.g., mass spectrometry). One exemplary use of arrays in the diagnostic methods of the invention is described below in more detail.

Use of a single gene in diagnostic applications. The diagnostic methods of the invention can focus on the expression of a single differentially expressed gene. For example, the diagnostic method can involve detecting a differentially expressed gene, or a polymorphism of such a gene (e.g., a polymorphism in a coding region or control region), that is associated with disease. Disease-associated polymorphisms can include deletion or truncation of the gene, mutations that alter expression level and/or affect activity of the encoded protein, etc.

A number of methods are available for analyzing nucleic acids for the presence of a specific sequence, e.g. a disease associated polymorphism. Where large amounts of DNA are available, genomic DNA is used directly. Alternatively, the region of interest is cloned into a suitable vector and grown in sufficient quantity for analysis. Cells that express a differentially expressed gene can be used as a source of mRNA, which can be assayed directly or reverse transcribed into cDNA for analysis. The nucleic acid can be amplified by conventional techniques, such as the polymerase chain reaction (PCR), to provide sufficient amounts for analysis, and a detectable label can be included in the amplification reaction (e.g., using a detectably labeled primer or detectably labeled oligonucleotides) to facilitate detection. Alternatively, various methods are also known in the art that utilize oligonucleotide ligation as a means of detecting polymorphisms, see, e.g., Riley et al., Nucl. Acids Res. (1990) 18:2887; and Delahunty et al., Am. J. Hum. Genet. (1996) 58:1239.

The amplified or cloned sample nucleic acid can be analyzed by one of a number of methods known in the art. The nucleic acid can be sequenced by dideoxy or other methods, and the sequence of bases compared to a selected sequence, e.g., to a wild-type sequence. Hybridization with the polymorphic or variant sequence can also be used to determine its presence in a sample (e.g., by Southern blot, dot blot, etc.). The hybridization pattern of a polymorphic or variant sequence and a control sequence to an array of oligonucleotide probes immobilized on a solid support, as described in US 5,445,934, or in WO 95/35505, can also be used as a means of identifying polymorphic or variant sequences associated with disease. Single strand conformational polymorphism (SSCP) analysis, denaturing gradient gel electrophoresis (DGGE), and heteroduplex analysis in gel matrices are used to detect conformational changes created by DNA sequence variation as alterations in electrophoretic mobility. Alternatively, where a polymorphism creates or destroys a recognition site for a restriction endonuclease, the sample is digested with that endonuclease, and the products size fractionated to determine whether the fragment was digested. Fractionation is performed by gel or capillary electrophoresis, particularly acrylamide or agarose gels.

Screening for mutations in a gene can be based on the functional or antigenic characteristics of the protein. Protein truncation assays are useful in detecting deletions that can affect the biological activity of the protein. Various immunoassays designed to detect polymorphisms in proteins can be used in screening. Where many diverse genetic mutations lead to a particular disease phenotype, functional protein assays have proven to be effective screening tools. The activity of the encoded protein can be determined by comparison with the wild-type protein.

Diagnosis, Prognosis, Assessment of Therapy (Therametrics), and Management of Cancer

The polynucleotides of the invention, as well as their gene products, are of particular interest as genetic or biochemical markers (e.g., in blood or tissues) that will detect the earliest changes along the carcinogenesis pathway and/or to monitor the efficacy of various therapies and preventive interventions. For example, the level of expression of certain polynucleotides can be indicative of a poorer prognosis, and therefore warrant more aggressive chemo- or radio-therapy for a patient or vice versa. The correlation of novel surrogate tumor specific features with response to treatment and outcome in patients can define prognostic indicators that allow the design of tailored therapy based on the molecular profile of the tumor. These therapies include antibody targeting, antagonists (e.g., small molecules), and gene therapy. Determining expression of certain polynucleotides and comparison of a patient's profile with known expression in normal tissue and variants of the disease allows a determination of the best possible treatment for a patient, both in terms of specificity of treatment and in terms of comfort level of the patient. Surrogate tumor markers, such as polynucleotide expression, can also be used to better classify, and thus diagnose and treat, different forms and disease states of cancer. Two classifications widely used in oncology that can benefit from identification of the expression levels of the genes corresponding to the polynucleotides of the invention are staging of the cancerous disorder, and grading the nature of the cancerous tissue.

The polynucleotides that correspond to differentially expressed genes, as well as their encoded gene products, can be useful to monitor patients having or susceptible to cancer to detect potentially malignant events at a molecular level before they are detectable at a gross morphological level. In addition, the polynucleotides of the invention, as well as the genes corresponding to such polynucleotides, can be useful as therametrics, e.g., to assess the effectiveness of therapy by using the polynucleotides or their encoded gene products, to assess, for example, tumor burden in the patient before, during, and after therapy.

Furthermore, a polynucleotide identified as corresponding to a gene that is differentially expressed in, and thus is important for, one type of cancer can also have implications for development or risk of development of other types of cancer, e.g., where a polynucleotide represents a gene differentially expressed across various cancer types. Thus, for example, expression of a polynucleotide corresponding to a gene that has clinical implications for metastatic colon cancer can also have clinical implications for stomach cancer or endometrial cancer.

Staging. Staging is a process used by physicians to describe how advanced the cancerous state is in a patient. Staging assists the physician in determining a prognosis, planning treatment and evaluating the results of such treatment. Staging systems vary with the types of cancer, but generally involve the following "TNM" system: the type of tumor, indicated by T; whether the cancer has metastasized to nearby lymph nodes, indicated by N; and whether the cancer has metastasized to more distant parts of the body, indicated by M. Generally, if a cancer is only detectable in the area of the primary lesion without having spread to any lymph nodes, it is called Stage I or Stage II, depending on the degree of invasiveness as indicated by the tumor grade of the primary lesion. If the primary lesion is of tumor grade I or II and the patient does not have any regional or distant metastasis, the cancer is classified as Stage I. If the primary lesion is of tumor grade III or IV and the patient does not have any regional or distant metastasis, the cancer is classified as Stage II. If the cancer has spread only to the regional lymph nodes, it is classified as Stage III. Cancers that have spread to a distant part of the body, such as liver, bone, brain or other sites, are Stage IV, the most advanced stage.

The polynucleotides of the invention can facilitate fine-tuning of the staging process by identifying markers for the aggresivity of a cancer, e.g., the metastatic potential, as well as the presence in different areas of the body. Thus, a Stage II cancer with a polynucleotide signifying a high metastatic potential cancer can be used to change a borderline Stage II tumor to a Stage III tumor, justifying more aggressive therapy. Conversely, the presence of a polynucleotide signifying a lower metastatic potential allows more conservative staging of a tumor.

Grading of cancers. Grade is a term used to describe how closely a tumor resembles normal tissue of its same type. The microscopic appearance of a tumor is used to identify tumor grade based on parameters such as cell morphology, cellular organization, and other markers of differentiation. As a general rule, the grade of a tumor corresponds to its rate of growth or aggressiveness, with undifferentiated or high-grade tumors being more aggressive than well-differentiated or low-grade tumors. The following guidelines are generally used for grading tumors: 1) GX Grade cannot be assessed; 2) G1 Well differentiated; 3) G2 Moderately well differentiated; 4) G3 Poorly differentiated; 5) G4 Undifferentiated. The polynucleotides of the invention can be especially valuable in determining the grade of the tumor, as they not only can aid in determining the differentiation status of the cells of a tumor, they can also identify factors other than differentiation that are valuable in determining the aggressiveness of a tumor, such as metastatic potential.

For prostate cancer, the Gleason Grading/Scoring system is most commonly used. A prostate biopsy tissue sample is examined under a microscope and a grade is assigned to the tissue based on: 1) the appearance of the cells, and 2) the arrangement of the cells. Each parameter is assessed on a scale of one (cells are almost normal) to five (abnormal), and the individual Gleason Grades are presented separated by a "+" sign. Alternatively, the two grades are combined to give a Gleason Score of 2-10. Thus, for a tissue sample that received a grade of 3 for each parameter, the Gleason Grade would be 3+3 and the Gleason Score would be 6. A lower Gleason Score indicates a well-differentiated tumor, while a higher Gleason Score indicates a poorly differentiated cancer that is more likely to spread. The majority of biopsies in general are Gleason Scores 5, 6 and 7.

| Gleason Score 2, 3, 4 | Gleason Score 5, 6, 7 | Gleason Score 8, 9, 10 |
|---|---|---|
| Low-grade tumor | Medium-grade tumor | High-grade tumor |
| Slow Growth | Unpredictable Growth | Aggressive Growth |
| Least dangerous.<br><br>Cells look most like normal prostate cells and are described as being "well-differentiated".<br><br>Tends to be slow growing. | Intermediate cancers may behave like low-grade or high-grade cancers.<br><br>The cells' behavior may depend on the volume of the cancer and the PSA level.<br><br>This is the most common grade of prostate cancer. | High-grade cancers are usually very aggressive and quick to spread to the tissue surrounding the prostate.<br><br>These cancer cells look least like normal prostate cells and are usually described as "poorly-differentiated". |

The polynucleotides of the Sequence Listing, and their corresponding genes and gene products, can be especially valuable in determining the grade of the tumor, as they not only can aid in determining the differentiation status of the cells of a tumor, they can also identify factors other than differentiation that are valuable in determining the aggressiveness of a tumor, such as metastatic potential.

Assessment of proliferation of cells in tumor. The differential expression level of the polynucleotides described herein can facilitate assessment of the rate of proliferation of tumor cells, and thus provide an indicator of the aggressiveness of the rate of tumor growth. For example, assessment of the relative expression levels of genes involved in the cell cycle can provide an indication of cellular proliferation, and thus serve as a marker of proliferation.

Detection of colon cancer. The polynucleotides corresponding to genes that exhibit the appropriate expression pattern can be used to detect colon cancer in a subject. Colorectal cancer is one of the most common neoplasms in humans and perhaps the most frequent form of hereditary neoplasia. Prevention and early detection are key factors in controlling and curing colorectal cancer. Colorectal cancer begins as polyps, which are small, benign growths of cells that form on the inner lining of the colon. Over a period of several years, some of these polyps accumulate additional mutations and become cancerous. Multiple familial colorectal cancer disorders have been identified, which are summarized as follows: 1) Familial adenomatous polyposis (FAP); 2) Gardner's syndrome; 3) Hereditary nonpolyposis colon cancer (HNPCC); and 4) Familial colorectal cancer in Ashkenazi Jews. The expression of appropriate polynucleotides of the invention can be used in the diagnosis, prognosis and management of colorectal cancer. Detection of colon cancer can be determined using expression levels of any of these sequences alone or in combination with the levels of expression. Determination of the aggressive nature and/or the metastatic potential of a colon cancer can be determined by comparing levels of one or more polynucleotides of the invention and comparing total levels of another sequence known to vary in cancerous tissue, e.g., expression of p53, DCC ras, lor FAP (see, e.g., Fearon ER, et al., Cell (1990) 61(5):759; Hamilton SR et al., Cancer (1993) 72:957;

Bodmer W, et al., Nat Genet. (1994) 4(3):217; Fearon ER, Ann N Y Acad Sci. (1995) 768:101). For example, development of colon cancer can be detected by examining the ratio of any of the polynucleotides of the invention to the levels of oncogenes (e.g., ras) or tumor suppressor genes (e.g., FAP or p53). Thus, expression of specific marker polynucleotides can be used to discriminate between normal and cancerous colon tissue, to discriminate between colon cancers with different cells of origin, to discriminate between colon cancers with different potential metastatic rates, etc. For a review of markers of cancer, see, e.g., Hanahan et al. (2000) Cell 100:57-70.

Detection of prostate cancer. The polynucleotides and their corresponding genes and gene products exhibiting the appropriate differential expression pattern can be used to detect prostate cancer in a subject. Prostate cancer is quite common in humans, with one out of every six men at a lifetime risk for prostate cancer, and can be relatively harmless or extremely aggressive. Some prostate tumors are slow growing, causing few clinical symptoms, while aggressive tumors spread rapidly to the lymph nodes, other organs and especially bone. Over 95% of primary prostate cancers are adenocarcinomas. Signs and symptoms may include: frequent urination, especially at night; inability to urinate; trouble starting or holding back urination; a weak or interrupted urine flow; and frequent pain or stiffness in the lower back, hips or upper thighs.

The prostate is divided into three areas - the peripheral zone, the transition zone, and the central zone - with a layer of tissue surrounding all three. Most prostate tumors form in the peripheral zone; the larger, glandular portion of the organ. Prostate cancer can also form in the tissue of the central zone. Surrounding the prostate is the prostate capsule, a tissue that separates the prostate from the rest of the body. When prostate cancer remains inside the prostate capsule, it is considered localized and treatable with surgery. Once the cancer punctures the capsule and spreads outside, treatment options are more limited. Prevention and early detection are key factors in controlling and curing prostate cancer.

While the Gleason Grade or Score of a prostate cancer can provide information useful in determining the appropriate treatment of a prostate cancer, the majority of prostate cancers are Gleason Scores 5, 6, and 7, which exhibit unpredictable behavior. These cancers may behave like less dangerous low-grade cancers or like extremely dangerous high-grade cancers. As a result, a patient living with a medium-grade prostate cancer is at constant risk of developing high-grade cancer.

The expression of appropriate polynucleotides can be used in the diagnosis, prognosis and management of prostate cancer. Detection of prostate cancer can be determined using expression levels of any of these sequences alone or in combination with the levels of expression of any other nucleotide sequences. Determination of the aggressive nature and/or the metastatic potential of a prostate cancer can be determined by comparing levels of one or more gene products of the genes corresponding to the polynucleotides described herein, and comparing total levels of another sequence known to vary in cancerous tissue, *e.g.*, expression of p53, DCC, ras, FAP (*see, e.g.*, Fearon ER, *et* *al.*, *Cell* (1990) *61(5)*:759; Hamilton SR *et al.*, *Cancer* (1993) *72*:957; Bodmer W, *et al.*, *Nat Genet.* (1994) *4(3)*:217; Fearon ER, *Ann N Y Acad Sci.* (1995) *768*:101).

For example, development of prostate cancer can be detected by examining the level of expression of a gene corresponding to a polynucleotides described herein to the levels of oncogenes (*e.g.* ras) or tumor suppressor genes (*e.g.* FAP or p53). Thus expression of specific marker polynucleotides can be used to discriminate between normal and cancerous prostate tissue, to discriminate between prostate cancers with different cells of origin, to discriminate between prostate cancers with different potential metastatic rates, etc. For a review of markers of cancer, see, *e.g.*, Hanahan *et al.* (2000) *Cell* 100:57-70.

In addition, many of the signs and symptoms of prostate cancer can be caused by a variety of other non-cancerous conditions. For example, one common cause of many of these signs and symptoms is a condition called benign prostatic hypertrophy, or BPH. In BPH, the prostate gets bigger and may block the flow of urine or interfere with sexual function. The methods and compositions of the invention can be used to distinguish between prostate cancer and such non-cancerous conditions. The methods of the invention can be used in conjunction with conventional methods of diagnosis, e.g., digital rectal exam and/or detection of the level of prostate specific antigen (PSA), a substance produced and secreted by the prostate, and/or prostatic acid phosphatase (PAP).

Detection of breast cancer. The majority of breast cancers are adenocarcinoma subtypes, which can be summarized as follows: 1) ductal carcinoma in situ (DCIS), including comedocarcinoma; 2) infiltrating (or invasive) ductal carcinoma (IDC); 3) lobular carcinoma in situ (LCIS); 4) infiltrating (or invasive) lobular carcinoma (ILC); 5) inflammatory breast cancer; 6) medullary carcinoma; 7) mucinous carcinoma; 8) Paget's disease of the nipple; 9) Phyllodes tumor; and 10) tubular carcinoma.

The expression of polynucleotides of the invention can be used in the diagnosis and management of breast cancer, as well as to distinguish between types of breast cancer. Detection of breast cancer can be determined using expression levels of any of the appropriate polynucleotides of the invention, either alone or in combination. Determination of the aggressive nature and/or the metastatic potential of a breast cancer can also be determined by comparing levels of one or more polynucleotides of the invention and comparing levels of another sequence known to vary in cancerous tissue, e.g., ER expression. In addition, development of breast cancer can be detected by examining the ratio of expression of a differentially expressed polynucleotide to the levels of steroid hormones (e.g., testosterone or estrogen) or to other hormones (e.g., growth hormone, insulin). Thus, expression of specific marker polynucleotides can be used to discriminate between normal and cancerous breast tissue, to discriminate between breast cancers with different cells of origin, to discriminate between breast cancers with different potential metastatic rates, etc.

Detection of lung cancer. The polynucleotides of the invention can be used to detect lung cancer in a subject. Although there are more than a dozen different kinds of lung cancer, the two main types of lung cancer are small cell and nonsmall cell, which encompass about 90% of all lung cancer cases. Small cell carcinoma (also called oat cell carcinoma) usually starts in one of the larger bronchial tubes, grows fairly rapidly, and is likely to be large by the time of diagnosis. Nonsmall cell lung cancer (NSCLC) is made up of three general subtypes of lung cancer. Epidermoid carcinoma (also called squamous cell carcinoma) usually starts in one of the larger bronchial tubes and grows relatively slowly. The size of these tumors can range from very small to quite large. Adenocarcinoma starts growing near the outside surface of the lung and can vary in both size and growth rate. Some slowly growing adenocarcinomas are described as alveolar cell cancer. Large cell carcinoma starts near the surface of the lung, grows rapidly, and the growth is usually fairly large when diagnosed. Other less common forms of lung cancer are carcinoid, cylindroma, mucoepidermoid, and malignant mesothelioma.

The polynucleotides of the invention, e.g., polynucleotides differentially expressed in normal cells versus cancerous lung cells (e.g., tumor cells of high or low metastatic potential) or between types of cancerous lung cells (e.g., high metastatic versus low metastatic), can be used to distinguish types of lung cancer as well as identifying traits specific to a certain patient's cancer and selecting an appropriate therapy. For example, if the patient's biopsy expresses a polynucleotide that is associated with a low metastatic potential, it may justify leaving a larger portion of the patient's lung in surgery to remove the lesion. Alternatively, a smaller lesion with expression of a polynucleotide that is associated with high metastatic potential may justify a more radical removal of lung tissue and/or the surrounding lymph nodes, even if no metastasis can be identified through pathological examination.

Tumor classification and patient stratification

The invention further provides for methods of classifying tumors, and thus grouping or "stratifying" patients, according to the expression profile of selected differentially expressed genes in a tumor. Differentially expressed genes can be analyzed for correlation with other differentially expressed genes in a single tumor type (*e.g.*, a prostate tumor) or between tumor types (*e.g.*, between prostate and colon tumors). Genes that demonstrate consistent correlation in expression profile in a given cancer cell type (*e.g.*, in a prostate cancer cell or type of prostate cancer) can be grouped together, *e.g.*, when one gene is overexpressed in a tumor, a second gene is also usually overexpressed. Tumors can then be classified according to the expression profile of one or more genes selected from one or more groups.

The tumor of each patient in a pool of potential patients can be classified as described above. Patients having similarly classified tumors can then be selected for participation in an investigative or clinical trial of a cancer therapeutic where a homogeneous population is desired. The tumor classification of a patient can also be used in assessing the efficacy of a cancer therapeutic in a heterogeneous patient population. In addition, therapy for a patient having a tumor of a given expression profile can then be selected accordingly.

Treatment of cancer

The invention further provides methods for reducing growth of cancer cells. In general, the methods comprise contacting a cancer cell with a substance that modulates (1) expression of a polynucleotide corresponding to a gene that is differentially expressed in cancer; or (2) a level of and/or an activity of a cancer-associated polypeptide. In general, the methods provide for decreasing the expression of a gene that is differentially expressed in a cancer cell (*e.g.,* overexpressed) or decreasing the level of and/or decreasing an activity of a cancer-associated polypeptide. The methods also provide for increasing expression of a gene that is underexpressed in a cancer cell or increasing the level of and/or increasing an activity of a cancer-associated polypeptide.

"Reducing growth of cancer cells" includes, but is not limited to, reducing proliferation of cancer cells (*e.g.,* prostate, colon, lung, breast, etc. cancer cells), and reducing the incidence of a non-cancerous cell becoming a cancerous cell. Whether a reduction in cancer cell growth has been achieved can be readily determined using any known assay, including, but not limited to, [$^3$H]-thymidine incorporation; counting cell number over a period of time; detecting and/or measuring a marker associated with the cancer type (e.g., CEA, CA19-9, LASA, PSA, PAP, CA15-3, CA27-29, NSE, LDH, etc.).

The present invention provides methods for treating cancer, generally comprising administering to an individual in need thereof a substance that reduces cancer cell growth, in an amount sufficient to reduce cancer cell growth and treat the cancer. Whether a substance, or a specific amount of the substance, is effective in treating cancer can be assessed using any of a variety of known diagnostic assays for the particular type of cancer being treated. The substance can be administered systemically or locally. Thus, in some embodiments, the substance is administered locally, and cancer growth is decreased at the site of administration. Local administration may be useful in treating, e.g., a solid tumor.

A substance that reduces cancer cell growth can be targeted to a cancer cell. Thus, in some embodiments, the invention provides a method of delivering a drug to a cancer cell, comprising administering a drug-antibody complex to a subject, wherein the antibody is specific for a particular cancer-associated polypeptide, and the drug is one that reduces cancer cell growth, a variety of which are known in the art. Targeting can be accomplished by coupling (e.g., linking, directly or via a linker molecule, either covalently or non-covalently, so as to form a drug-antibody complex) a drug to an antibody specific for a particular cancer-associated polypeptide. Methods of coupling a drug to an antibody are well known in the art and need not be elaborated upon herein.

In another embodiment, differentially expressed gene products (*e.g.*, polypeptides or polynucleotides encoding such polypeptides) may be effectively used in treatment through vaccination. The growth of cancer cells is naturally limited in part due to immune surveillance. Stimulation of the immune system using a particular tumor-specific antigen enhances the effect towards the tumor expressing the antigen. An active vaccine comprising a polypeptide encoded by the cDNA of this invention would be appropriately administered to subjects having overabundance of the corresponding RNA, or those predisposed for developing cancer cells with overabundance of the same RNA. Polypeptide antigens are typically combined with an adjuvant as part of a vaccine composition. The vaccine is preferably administered first as a priming dose, and then again as a boosting dose, usually at least four weeks later. Further boosting doses may be given to enhance the effect. The dose and its timing are usually determined by the person responsible for the treatment.

The invention also encompasses the selection of a therapeutic regimen based upon the expression profile of differentially expressed genes in the patient's tumor. For example, a tumor can be analyzed for its expression profile of the genes corresponding to SEQ ID NOS:1-1542 as described herein, *e.g.*, the tumor is analyzed to determine which genes are expressed at elevated levels or at decreased levels relative to normal cells of the same tissue type. The expression patterns of the tumor are then compared to the expression patterns of tumors that respond to a selected therapy. Where the expression profiles of the test tumor cell and the expression profile of a tumor cell of known drug responsivity at least substantially match (*e.g.*, selected sets of genes at elevated levels in the tumor of known drug responsivitiy and are also at elevated levels in the test tumor cell), then the drug selected for therapy is the drug to which tumors with that expression pattern respond.

Identification of Therapeutic Targets and Anti-Cancer Therapeutic Agents

The present invention also encompasses methods for identification of agents having the ability to modulate activity of a differentially expressed gene product, as well as methods for identifying a differentially expressed gene product as a therapeutic target for treatment of cancer, especially prostate cancer.

Candidate agents

Identification of compounds that modulate activity of a differentially expressed gene product can be accomplished using any of a variety of drug screening techniques. Such agents are candidates for development of cancer therapies. Of particular interest are screening assays for agents that have tolerable toxicity for normal, non-cancerous human cells. The screening assays of the invention are generally based upon the ability of the agent to modulate an activity of a differentially expressed gene product and/or to inhibit or suppress phenomenon associated with cancer (*e.g.*, cell proliferation, colony formation, cell cycle arrest, metastasis, and the like).

The term "agent" as used herein describes any molecule, *e.g.* protein or pharmaceutical, with the capability of modulating a biological activity of a gene product of a differentially expressed gene.

Generally a plurality of assay mixtures are run in parallel with different agent concentrations to obtain a differential response to the various concentrations. Typically, one of these concentrations serves as a negative control, *i.e.* at zero concentration or below the level of detection.

Candidate agents encompass numerous chemical classes, though typically they are organic molecules, preferably small organic compounds having a molecular weight of more than 50 and less than about 2,500 daltons. Candidate agents comprise functional groups necessary for structural interaction with proteins, particularly hydrogen bonding, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemical groups. The candidate agents often comprise cyclical carbon or heterocyclic structures and/or aromatic or polyaromatic structures substituted with one or more of the above functional groups. Candidate agents are also found among biomolecules including, but not limited to: peptides, saccharides, fatty acids, steroids, purines, pyrimidines, derivatives, structural analogs or combinations thereof.

Candidate agents are obtained from a wide variety of sources including libraries of synthetic or natural compounds. For example, numerous means are available for random and directed synthesis of a wide variety of organic compounds and biomolecules, including expression of randomized oligonucleotides and oligopeptides. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts (including extracts from human tissue to identify endogenous factors affecting differentially expressed gene products) are available or readily produced. Additionally, natural or synthetically produced libraries and compounds are readily modified through conventional chemical, physical and biochemical means, and may be used to produce combinatorial libraries. Known pharmacological agents may be subjected to directed or random chemical modifications, such as acylation, alkylation, esterification, amidification, *etc.* to produce structural analogs.

Exemplary candidate agents of particular interest include, but are not limited to, antisense polynucleotides, and antibodies, soluble receptors, and the like. Antibodies and soluble receptors are of particular interest as candidate agents where the target differentially expressed gene product is secreted or accessible at the cell-surface (*e.g.*, receptors and other molecule stably-associated with the outer cell membrane).

<u>Screening of candidate agents</u>

Screening assays can be based upon any of a variety of techniques readily available and known to one of ordinary skill in the art. In general, the screening assays involve contacting a cancerous cell (preferably a cancerous prostate cell) with a candidate agent, and assessing the effect upon biological activity of a differentially expressed gene product. The effect upon a biological activity can be detected by, for example, detection of expression of a gene product of a differentially expressed gene (*e.g.*, a decrease in mRNA or polypeptide levels, would in turn cause a decrease in biological activity of the gene product). Alternatively or in addition, the effect of the candidate agent can be assessed by examining the effect of the candidate agent in a functional assay. For example, where the differentially expressed gene product is an enzyme, then the effect upon biological activity can be assessed by detecting a level of enzymatic activity associated with the differentially expressed gene product. The functional assay will be selected according to the differentially expressed gene product. In general, where the differentially expressed gene is increased in expression in a cancerous cell, agents of interest are those that decrease activity of the differentially expressed gene product.

Assays described infra can be readily adapted in the screening assay embodiments of the invention. Exemplary assays useful in screening candidate agents include, but are not limited to, hybridization-based assays (*e.g.*, use of nucleic acid probes or primers to assess expression levels), antibody-based assays (*e.g.*, to assess levels of polypeptide gene products), binding assays (*e.g.*, to detect interaction of a candidate agent with a differentially expressed polypeptide, which assays may be competitive assays where a natural or synthetic ligand for the polypeptide is available), and the like. Additional exemplary assays include, but are not necessarily limited to, cell proliferation assays, antisense knockout assays, assays to detect inhibition of cell cycle, assays of induction of cell death/apoptosis, and the like. Generally such assays are conducted *in vitro*, but many assays can be adapted for *in vivo* analyses, *e.g.*, in an animal model of the cancer.

Identification of therapeutic targets

In another embodiment, the invention contemplates identification of differentially expressed genes and gene products as therapeutic targets. In some respects, this is the converse of the assays described above for identification of agents having activity in modulating (*e.g.*, decreasing or increasing) activity of a differentially expressed gene product.

In this embodiment, therapeutic targets are identified by examining the effect(s) of an agent that can be demonstrated or has been demonstrated to modulate a cancerous phenotype (*e.g.*, inhibit or suppress or prevent development of a cancerous phenotype). Such agents are generally referred to herein as an "anti-cancer agent", which agents encompass chemotherapeutic agents. For example, the agent can be an antisense oligonucleotide that is specific for a selected gene transcript. For example, the antisense oligonucleotide may have a sequence corresponding to a sequence of a differentially expressed gene described herein, *e.g.*, a sequence of one of SEQ ID NOS:1-2164.

Assays for identification of therapeutic targets can be conducted in a variety of ways using methods that are well known to one of ordinary skill in the art. For example, a test cancerous cell that expresses or overexpresses a differentially expressed gene is contacted with an anti-cancer agent, the effect upon a cancerous phenotype and a biological activity of the candidate gene product assessed. The biological activity of the candidate gene product can be assayed be examining, for example, modulation of expression of a gene encoding the candidate gene product (*e.g.*, as detected by, for example, an increase or decrease in transcript levels or polypeptide levels), or modulation of an enzymatic or other activity of the gene product. The cancerous phenotype can be, for example, cellular proliferation, loss of contact inhibition of growth (*e.g.*, colony formation), tumor growth (*in vitro* or *in vivo*), and the like. Alternatively or in addition, the effect of modulation of a biological activity of the candidate target gene upon cell death/apoptosis or cell cycle regulation can be assessed.

Inhibition or suppression of a cancerous phenotype, or an increase in cell/death apoptosis as a result of modulation of biological activity of a candidate gene product indicates that the candidate gene product is a suitable target for cancer therapy. Assays described infra can be readily adapted in for assays for identification of therapeutic targets. Generally such assays are conducted *in vitro*, but many assays can be adapted for *in vivo* analyses, *e.g.*, in an appropriate, art-accepted animal model of the cancer.

Use of Polynucleotides to Screen for Peptide Analogs and Antagonists

Polypeptides encoded by the instant polynucleotides and corresponding full-length genes can be used to screen peptide libraries to identify binding partners, such as receptors, from among the encoded polypeptides. Peptide libraries can be synthesized according to methods known in the art (see, e.g., USPN 5,010,175 , and WO 91/17823).

Agonists or antagonists of the polypeptides of the invention can be screened using any available method known in the art, such as signal transduction, antibody binding, receptor binding, mitogenic assays, chemotaxis assays, etc. The assay conditions ideally should resemble the conditions under which the native activity is exhibited in vivo, that is, under physiologic pH, temperature, and ionic strength. Suitable agonists or antagonists will exhibit strong inhibition or enhancement of the native activity at concentrations that do not cause toxic side effects in the subject. Agonists or antagonists that compete for binding to the native polypeptide can require concentrations equal to or greater than the native concentration, while inhibitors capable of binding irreversibly to the polypeptide can be added in concentrations on the order of the native concentration.

Such screening and experimentation can lead to identification of a novel polypeptide binding partner, such as a receptor, encoded by a gene or a cDNA corresponding to a polynucleotide of the invention, and at least one peptide agonist or antagonist of the novel binding partner. Such agonists and antagonists can be used to modulate, enhance, or inhibit receptor function in cells to which the receptor is native, or in cells that possess the receptor as a result of genetic engineering. Further, if the novel receptor shares biologically important characteristics with a known receptor, information about agonist/antagonist binding can facilitate development of improved agonists/antagonists of the known receptor.

Vaccines and Uses

The differentially expressed nucleic acids and polypeptides produced by the nucleic acids of the invention can also be used to modulate primary immune response to prevent or treat cancer. Every immune response is a complex and intricately regulated sequence of events involving several cell types. It is triggered when an antigen enters the body and encounters a specialized class of cells called antigen-presenting cells (APCs). These APCs capture a minute amount of the antigen and display it in a form that can be recognized by antigen-specific helper T lymphocytes. The helper (Th) cells become activated and, in turn, promote the activation of other classes of lymphocytes, such as B cells or cytotoxic T cells. The activated lymphocytes then proliferate and carry out their specific effector functions, which in many cases successfully activate or eliminate the antigen. Thus, activating the immune response to a particular antigen associated with a cancer cell can protect the patient from developing cancer or result in lymphocytes eliminating cancer cells expressing the antigen.

Gene products, including polypeptides, mRNA (particularly mRNAs having distinct secondary and/or tertiary structures), cDNA, or complete gene, can be prepared and used in vaccines for the treatment or prevention of hyperproliferative disorders and cancers. The nucleic acids and polypeptides can be utilized to enhance the immune response, prevent tumor progression, prevent hyperproliferative cell growth, and the like. Methods for selecting nucleic acids and polypeptides that are capable of enhancing the immune response are known in the art. Preferably, the gene products for use in a vaccine are gene products which are present on the surface of a cell and are recognizable by lymphocytes and antibodies.

The gene products may be formulated with pharmaceutically acceptable carriers into pharmaceutical compositions by methods known in the art. The composition is useful as a vaccine to prevent or treat cancer. The composition may further comprise at least one co-immunostimulatory molecule, including but not limited to one or more major histocompatibility complex (MHC) molecules, such as a class I or class II molecule, preferably a class I molecule. The composition may further comprise other stimulator molecules including B7.1, B7.2, ICAM-1, ICAM-2, LFA-1, LFA-3, CD72 and the like, immunostimulatory polynucleotides (which comprise an 5'-CG-3' wherein the cytosine is unmethylated), and cytokines which include but are not limited to IL-1 through IL-15, TNF-α, IFN-γ, RANTES, G-CSF, M-CSF, IFN-α, CTAP III, ENA-78, GRO, I-309, PF-4, IP-10, LD-78, MGSA, MIP-1α, MIP-1β, or combination thereof, and the like for immunopotentiation. In one embodiment, the immunopotentiators of particular interest are those which facilitate a Th1 immune response.

The gene products may also be prepared with a carrier that will protect the gene products against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, polylactic acid, and the like. Methods for preparation of such formulations are known in the art.

In the methods of preventing or treating cancer, the gene products may be administered via one of several routes including but not limited to transdermal, transmucosal, intravenous, intramuscular, subcutaneous, intradermal, intraperitoneal, intrathecal, intrapleural, intrauterine, rectal, vaginal, topical, intratumor, and the like. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, administration bile salts and fusidic acid derivatives. In addition, detergents may be used to facilitate permeation. Transmucosal administration may be by nasal sprays or suppositories. For oral administration, the gene products are formulated into conventional oral administration form such as capsules, tablets and toxics.

The gene product is administered to a patient in an amount effective to prevent or treat cancer. In general, it is desirable to provide the patient with a dosage of gene product of at least about 1 pg per Kg body weight, preferably at least about 1 ng per Kg body weight, more preferably at least about 1 µg or greater per Kg body weight of the recipient. A range of from about 1 ng per Kg body weight to about 100 mg per Kg body weight is preferred although a lower or higher dose may be administered. The dose is effective to prime, stimulate and/or cause the clonal expansion of antigen-specific T lymphocytes, preferably cytotoxic T lymphocytes, which in turn are capable of preventing or treating cancer in the recipient. The dose is administered at least once and may be provided as a bolus or a continuous administration. Multiple administrations of the dose over a period of several weeks to months may be preferable. Subsequent doses may be administered as indicated.

In another method of treatment, autologous cytotoxic lymphocytes or tumor infiltrating lymphocytes may be obtained from a patient with cancer. The lymphocytes are grown in culture, and antigen-specific lymphocytes are expanded by culturing in the presence of the specific gene products alone or in combination with at least one co-immunostimulatory molecule with cytokines. The antigen-specific lymphocytes are then infused back into the patient in an amount effective to reduce or eliminate the tumors in the patient. Cancer vaccines and their uses are further described in USPN 5,961,978; USPN 5,993,829; USPN 6,132,980; and WO 00/38706.

Pharmaceutical Compositions and Uses

Pharmaceutical compositions can comprise polypeptides, receptors that specifically bind a polypeptide produced by a differentially expressed gene (e.g., antibodies, or polynucleotides (including antisense nucleotides and ribozymes) of the claimed invention in a therapeutically effective amount. The compositions can be used to treat primary tumors as well as metastases of primary tumors. In addition, the pharmaceutical compositions can be used in conjunction with conventional methods of cancer treatment, e.g., to sensitize tumors to radiation or conventional chemotherapy.

Where the pharmaceutical composition comprises a receptor (such as an antibody) that specifically binds to a gene product encoded by a differentially expressed gene, the receptor can be coupled to a drug for delivery to a treatment site or coupled to a detectable label to facilitate imaging of a site comprising colon cancer cells. Methods for coupling antibodies to drugs and detectable labels are well known in the art, as are methods for imaging using detectable labels.

The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature.

The precise effective amount for a subject will depend upon the subject's size and health, the nature and extent of the condition, and the therapeutics or combination of therapeutics selected for administration. Thus, it is not useful to specify an exact effective amount in advance. However, the effective amount for a given situation is determined by routine experimentation and is within the judgment of the clinician. For purposes of the present invention, an effective dose will generally be from about 0.01 mg/kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to any pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which can be administered without undue toxicity. Suitable carriers can be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Pharmaceutically acceptable carriers in therapeutic compositions can include liquids such as water, saline, glycerol and ethanol. Auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, can also be present in such vehicles.

Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection can also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier. Pharmaceutically acceptable salts can also be present in the pharmaceutical composition, e.g., mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable excipients is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

Delivery Methods

Once formulated, the compositions of the invention can be (1) administered directly to the subject (e.g., as polynucleotide or polypeptides); or (2) delivered ex vivo, to cells derived from the subject (e.g., as in ex vivo gene therapy). Direct delivery of the compositions will generally be accomplished by parenteral injection, e.g., subcutaneously, intraperitoneally, intravenously or intramuscularly, intratumorally or to the interstitial space of a tissue. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal applications, needles, and gene guns or hyposprays. Dosage treatment can be a single dose schedule or a multiple dose schedule.

Methods for the ex vivo delivery and reimplantation of transformed cells into a subject are known in the art and described in, e.g., WO 93/14778. Examples of cells useful in ex vivo applications include, for example, stem cells, particularly hematopoetic, lymph cells, macrophages, dendritic cells, or tumor cells. Generally, delivery of nucleic acids for both ex vivo and in vitro applications can be accomplished by, for example, dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei, all well known in the art.

Once differential expression of a gene corresponding to a polynucleotide of the invention has been found to correlate with a proliferative disorder, such as neoplasia, dysplasia, and hyperplasia, the disorder can be amenable to treatment by administration of a therapeutic agent based on the provided polynucleotide, corresponding polypeptide or other corresponding molecule (e.g., antisense, ribozyme, etc.). In other embodiments, the disorder can be amenable to treatment by administration of a small molecule drug that, for example, serves as an inhibitor (antagonist) of the function of the encoded gene product of a gene having increased expression in cancerous cells relative to normal cells or as an agonist for gene products that are decreased in expression in cancerous cells (e.g., to promote the activity of gene products that act as tumor suppressors).

The dose and the means of administration of the inventive pharmaceutical compositions are determined based on the specific qualities of the therapeutic composition, the condition, age, and weight of the patient, the progression of the disease, and other relevant factors. For example, administration of polynucleotide therapeutic composition agents of the invention includes local or systemic administration, including injection, oral administration, particle gun or catheterized administration, and topical administration. Preferably, the therapeutic polynucleotide composition contains an expression construct comprising a promoter operably linked to a polynucleotide of at least 12, 22, 25, 30, or 35 contiguous nt of the polynucleotide of the invention. Various methods can be used to administer the therapeutic composition directly to a specific site in the body. For example, a small metastatic lesion is located and the therapeutic composition injected several times in several different locations within the body of tumor. Alternatively, arteries that serve a tumor are identified, and the therapeutic composition injected into such an artery, in order to deliver the composition directly into the tumor. A tumor that has a necrotic center is aspirated and the composition injected directly into the now empty center of the tumor. The antisense composition is directly administered to the surface of the tumor, for example, by topical application of the composition. X-ray imaging is used to assist in certain of the above delivery methods.

Targeted delivery of therapeutic compositions containing an antisense polynucleotide, subgenomic polynucleotides, or antibodies to specific tissues can also be used. Receptor-mediated DNA delivery techniques are described in, for example, Findeis et al., Trends Biotechnol. (1993) 11:202; Chiou et al., Gene Therapeutics: Methods And Applications Of Direct Gene Transfer (J.A. Wolff, ed.) (1994); Wu et al., J. Biol. Chem. (1988) 263:621; Wu et al., J. Biol. Chem. (1994) 269:542; Zenke et al., Proc. Natl. Acad. Sci. (USA) (1990) 87:3655; Wu et al., J. Biol. Chem. (1991) 266:338. Therapeutic compositions containing a polynucleotide are administered in a range of about 100 ng to about 200 mg of DNA for local administration in a gene therapy protocol. Concentration ranges of about 500 ng to about 50 mg, about 1 micrograms to about 2 mg, about 5 micrograms to about 500 micrograms, and about 20 micrograms to about 100 micrograms of DNA can also be used during a gene therapy protocol. Factors such as method of action (e.g., for enhancing or inhibiting levels of the encoded gene product) and efficacy of transformation and expression are considerations which will affect the dosage required for ultimate efficacy of the antisense subgenomic polynucleotides.

Where greater expression is desired over a larger area of tissue, larger amounts of antisense subgenomic polynucleotides or the same amounts readministered in a successive protocol of administrations, or several administrations to different adjacent or close tissue portions of, for example, a tumor site, may be required to effect a positive therapeutic outcome. In all cases, routine experimentation in clinical trials will determine specific ranges for optimal therapeutic effect. For polynucleotide related genes encoding polypeptides or proteins with anti-inflammatory activity, suitable use, doses, and administration are described in USPN 5,654,173.

The therapeutic polynucleotides and polypeptides of the present invention can be delivered using gene delivery vehicles. The gene delivery vehicle can be of viral or non-viral origin (see generally, Jolly, Cancer Gene Therapy (1994) 1:51; Kimura, Human Gene Therapy (1994) 5:845; Connelly, Human Gene Therapy (1995) 1:185; and Kaplitt, Nature Genetics (1994) 6:148). Expression of such coding sequences can be induced using endogenous mammalian or heterologous promoters. Expression of the coding sequence can be either constitutive or regulated.

Viral-based vectors for delivery of a desired polynucleotide and expression in a desired cell are well known in the art. Exemplary viral-based vehicles include, but are not limited to, recombinant retroviruses (see, e.g., WO 90/07936; WO 94/03622; WO 93/25698; WO 93/25234; USPN 5, 219,740; WO 93/11230; WO 93/10218; USPN 4,777,127; GB Patent No. 2,200,651; EP 0 345 242; and WO 91/02805), alphavirus-based vectors (e.g., Sindbis virus vectors, Semliki forest virus (ATCC VR-67; ATCC VR-1247), Ross River virus (ATCC VR-373; ATCC VR-1246) and Venezuelan equine encephalitis virus (ATCC VR-923; ATCC VR-1250; ATCC VR 1249; ATCC VR-532), and adeno-associated virus (AAV) vectors (see, e.g., WO 94/12649, WO 93/03769; WO 93/19191; WO

94/28938; WO 95/11984 and WO 95/00655). Administration of DNA linked to killed adenovirus, as described in Curiel, Hum. Gene Ther. (1992) 3:147, can also be employed.

Non-viral delivery vehicles and methods can also be employed, including, but not limited to, polycationic condensed DNA linked or unlinked to killed adenovirus alone (see, e.g., Curiel, Hum. Gene Ther. (1992) 3:147); ligand-linked DNA (see, e.g., Wu, J. Biol. Chem. (1989) 264:16985); eukaryotic cell delivery vehicles cells (see, e.g., USPN 5,814,482; WO 95/07994; WO 96/17072; WO 95/30763; and WO 97/42338) and nucleic charge neutralization or fusion with cell membranes. Naked DNA can also be employed. Exemplary naked DNA introduction methods are described in WO 90/11092 and USPN 5,580,859. Liposomes that can act as gene delivery vehicles are described in USPN 5,422,120; WO 95/13796; WO 94/23697; WO 91/14445; and EP 0524968. Additional approaches are described in Philip, Mol. Cell Biol. (1994) 14:2411, and in Woffendin, Proc. Natl. Acad. Sci. (1994) 91:1581

Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in Woffendin et al., Proc. Natl. Acad. Sci. USA (1994) 91(24):11581. Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials or use of ionizing radiation (see, e.g., USPN 5,206,152 and WO 92/11033). Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for example, use of hand-held gene transfer particle gun (see, e.g., USPN 5,149,655); use of ionizing radiation for activating transferred gene (see, e.g., USPN 5,206,152 and WO 92/11033).

The present invention will now be illustrated by reference to the following examples which set forth particularly advantageous embodiments. However, it should be noted that these embodiments are illustrative and are not to be construed as restricting the invention in any way.

## EXAMPLES

The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the present invention, and are not intended to limit the scope of what the inventors regard as their invention nor are they intended to represent that the experiments below are all or the only experiments performed. It will be readily apparent to those skilled in the art that the formulations, dosages, methods of administration, and other parameters of this invention may be further modified or substituted in various ways without departing from the spirit and scope of the invention. Efforts have been made to ensure accuracy with respect to numbers used (e.g. amounts, temperature, etc.) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, molecular weight is weight average molecular weight, temperature is in degrees Centigrade, and pressure is at or near atmospheric.

Example 1:Source of Biological Materials and Overview of Novel Polynucleotides Expressed by the Biological Materials

Candidate polynucleotides that may represent novel polynucleotides were obtained from cDNA libraries generated from selected cell lines and patient tissues. In order to obtain the candidate polynucleotides, mRNA was isolated from several selected cell lines and patient tissues, and used to construct cDNA libraries. The cells and tissues that served as sources for these cDNA libraries are summarized in Table 1 below.

Human colon cancer cell line Km12L4-A (Morikawa, et al., Cancer Research (1988) 48:6863) is derived from the KM12C cell line. The KM12C cell line (Morikawa et al. Cancer Res. (1988) 48:1943-1948), which is poorly metastatic (low metastatic) was established in culture from a Dukes' stage B2 surgical specimen (Morikawa et al. Cancer Res. (1988) 48:6863). The KM12L4-A is a highly metastatic subline derived from KM12C (Yeatman et al. Nucl. Acids. Res. (1995) 23:4007; Bao-Ling et al. Proc. Annu. Meet. Am. Assoc. Cancer. Res. (1995) 21:3269). The KM12C and KM12C-derived cell lines (e.g., KM12L4, KM12L4-A, etc.) are well-recognized in the art as a model cell line for the study of colon cancer (see, e.g., Moriakawa et al., supra; Radinsky et al. Clin. Cancer Res. (1995) 1:19; Yeatman et al., (1995) supra; Yeatman et al. Clin. Exp. Metastasis (1996) 14:246).

The MDA-MB-231 cell line (Brinkley et al. Cancer Res. (1980) 40:3118-3129) was originally isolated from pleural effusions (Cailleau, J. Natl. Cancer. Inst. (1974) 53:661), is of high metastatic potential, and forms poorly differentiated adenocarcinoma grade II in nude mice consistent with breast carcinoma. The MCF7 cell line was derived from a pleural effusion of a breast adenocarcinoma and is non-metastatic. The MV-522 cell line is derived from a human lung carcinoma and is of high metastatic potential. The UCP-3 cell line is a low metastatic human lung carcinoma cell line; the MV-522 is a high metastatic variant of UCP-3. These cell lines are well-recognized in the art as models for the study of human breast and lung cancer (see, e.g., Chandrasekaran et al., Cancer Res. (1979) 39:870 (MDA-MB-231 and MCF-7); Gastpar et al., J Med Chem (1998) 41:4965 (MDA-MB-231 and MCF-7); Ranson et al., Br J Cancer (1998) 77:1586 (MDA-MB-231 and MCF-7); Kuang et al., Nucleic Acids Res (1998) 26:1116 (MDA-MB-231 and MCF-7); Varki et al., Int J Cancer (1987) 40:46 (UCP-3); Varki et al., Tumour Biol. (1990) 11:327; (MV-522 and UCP-3); Varki et al., Anticancer Res. (1990) 10:637; (MV-522); Kelner et al., Anticancer Res (1995) 15:867 (MV-522); and Zhang et al., Anticancer Drugs (1997) 8:696 (MV522)).

The samples of libraries 15-20 are derived from two different patients (UC#2, and UC#3). The bFGF-treated HMVEC were prepared by incubation with bFGF at 10ng/ml for 2 hrs; the VEGF-treated HMVEC were prepared by incubation with 20ng/ml VEGF for 2 hrs. Following incubation with the respective growth factor, the cells were washed and lysis buffer added for RNA preparation.

GRRpz was derived from normal prostate epithelium. The WOca cell line is a Gleason Grade 4 cell line.

The source materials for generating the normalized prostate libraries of libraries 25 and 26 were cryopreserved prostate tumor tissue from a patient with Gleason grade 3+3 adenocarcinoma and matched normal prostate biopsies from a pool of at-risk subjects under medical surveillance. The source materials for generating the normalized prostate libraries of libraries 30 and 31 were cryopreserved prostate tumor tissue from a patient with Gleason grade 4+4 adenocarcinoma and matched normal prostate biopsies from a pool of at-risk subjects under medical surveillance.

The source materials for generating the normalized breast libraries of libraries 27, 28 and 29 were cryopreserved breast tissue from a primary breast tumor (infiltrating ductal carcinoma)(library 28), from a lymph node metastasis (library 29), or matched normal breast biopsies from a pool of at-risk subjects under medical surveillance. In each case, prostate or breast epithelia were harvested directly from frozen sections of tissue by laser capture microdissection (LCM, Arcturus Enginering Inc., Mountain View, CA), carried out according to methods well known in the art (*see*, Simone et al. Am J Pathol. 156(2):445-52 (2000)), to provide substantially homogenous cell samples.

**Table 1.** Description of cDNA Libraries

| Library (lib#) | Description | Number of Clones in Library |
|---|---|---|
| 0 | Artificial library composed of deselected clones (clones with no associated variant or cluster) | 673 |
| 1 | Human Colon Cell Line Km12 L4: High Metastatic Potential (derived from Km12C) | 308731 |
| 2 | Human Colon Cell Line Km12C: Low Metastatic Potential | 284771 |
| 3 | Human Breast Cancer Cell Line MDA-MB-231: High Metastatic Potential; micro-mets in lung | 326937 |
| 4 | Human Breast Cancer Cell Line MCF7: Non Metastatic | 318979 |
| 8 | Human Lung Cancer Cell Line MV-522: High Metastatic Potential | 223620 |
| 9 | Human Lung Cancer Cell Line UCP-3: Low Metastatic Potential | 312503 |
| 12 | Human microvascular endothelial cells (HMEC) - UNTREATED (PCR (OligodT) cDNA library) | 41938 |
| 13 | Human microvascular endothelial cells (HMEC) - bFGF TREATED (PCR (OligodT) cDNA library) | 42100 |
| 14 | Human microvascular endothelial cells (HMEC) - VEGF TREATED (PCR (OligodT) cDNA library) | 42825 |
| 15 | Normal Colon - UC#2 Patient (MICRODISSECTED PCR (OligodT) cDNA library) | 282722 |
| 16 | Colon Tumor - UC#2 Patient (MICRODISSECTED PCR (OligodT) cDNA library) | 298831 |
| 17 | Liver Metastasis from Colon Tumor of UC#2 Patient (MICRODISSECTED PCR (OligodT) cDNA library) | 303467 |
| 18 | Normal Colon - UC#3 Patient (MICRODISSECTED PCR (OligodT) cDNA library) | 36216 |
| 19 | Colon Tumor - UC#3 Patient (MICRODISSECTED PCR (OligodT) cDNA library) | 41388 |
| 20 | Liver Metastasis from Colon Tumor of UC#3 Patient (MICRODISSECTED PCR (OligodT) cDNA library) | 30956 |

| Library (lib#) | Description | Number of Clones in Library |
|---|---|---|
| 21 | GRRpz Cells derived from normal prostate epithelium | 164801 |
| 22 | WOca Cells derived from Gleason Grade 4 prostate cancer epithelium | 162088 |
| 23 | Normal Lung Epithelium of Patient #1006 (MICRODISSECTED PCR (OligodT) cDNA library) | 306198 |
| 24 | Primary tumor, Large Cell Carcinoma of Patient #1006 (MICRODISSECTED PCR (OligodT) cDNA library) | 309349 |
| 25 | Normal Prostate Epithelium from Patient IF97-26811 | 279444 |
| 26 | Prostate Cancer Epithelium Gleason 3+3 Patient IF97-26811 | 269406 |
| 27 | Normal Breast Epithelium from Patient 515 | 239494 |
| 28 | Primary Breast tumor from Patient 515 | 259960 |
| 29 | Lymph node metastasis from Patient 515 | 326786 |
| 30 | Normal Prostate Epithelium from Chiron Patient ID 884 | 298431 |
| 31 | Prostate Cancer Epithelium (Gleason 4+4) from Chiron Patient ID 884 | 331941 |

Characterization of sequences in the libraries

After using the software program Phred (ver 0.000925.c, Green and Weing,, ©1993-2000) to select those polynucleotides having the best quality sequence, the polynucleotides were compared against the public databases to identify any homologous sequences. The sequences of the isolated polynucleotides were first masked to eliminate low complexity sequences using the RepeatMasker masking program, publicly available through a web site supported by the University of Washington (*See also* Smit, A.F.A. and Green, P., unpublished results). Generally, masking does not influence the final search results, except to eliminate sequences of relatively little interest due to their low complexity, and to eliminate multiple "hits" based on similarity to repetitive regions common to multiple sequences, e.g., Alu repeats.

The remaining sequences were then used in a homology search of the GenBank database using the TeraBLAST program (TimeLogic, Crystal Bay, Nevada). TeraBLAST is a version of the publicly available BLAST search algorithm developed by the National Center for Biotechnology, modified to operate at an accelerated speed with increased sensitivity on a specialized computer hardware platform. The program was run with the default parameters recommended by TimeLogic to provide the best sensitivity and speed for searching DNA and protein sequences. Sequences that exhibited greater than 70% overlap, 99% identity, and a p value of less than 1 x 10e-40 were discarded. Sequences from this search also were discarded if the inclusive parameters were met, but the sequence was ribosomal or vector-derived.

The resulting sequences from the previous search were classified into three groups (1, 2 and 3 below) and searched in a TeraBLASTX vs. NRP (non-redundant proteins) database search: (1) unknown (no hits in the GenBank search), (2) weak similarity (greater than 45% identity and p value of less than 1 x 10e-5), and (3) high similarity (greater than 60% overlap, greater than 80% identity, and p value less than 1 x 10e-5). Sequences having greater than 70% overlap, greater than 99% identity, and p value of less than 1 x 10e-40 were discarded.

The remaining sequences were classified as unknown (no hits), weak similarity, and high similarity (parameters as above). Two searches were performed on these sequences. First, a TeraBLAST vs. EST database search was performed and sequences with greater than 99% overlap, greater than 99% similarity and a p value of less than 1 x 10e-40 were discarded. Sequences with a p value of less than 1 x 10e-65 when compared to a database sequence of human origin were also excluded. Second, a TeraBLASTN vs. Patent GeneSeq database was performed and sequences having greater than 99% identity, p value less than 1 x 10e-40, and greater than 99% overlap were discarded.

The remaining sequences were subjected to screening using other rules and redundancies in the dataset. Sequences with a p value of less than 1 x 10e-111 in relation to a database sequence of human origin were specifically excluded. The final result provided the sequences listed as SEQ ID NOS:1-1219 in the accompanying Sequence Listing and summarized in Table 2 (inserted prior to claims). Each identified polynucleotide represents sequence from at least a partial mRNA transcript.

Summary of polynucleotides of the invention

Table 2 (inserted prior to claims) provides a summary of polynucleotides isolated as described. Specifically, Table 2 provides: 1) the SEQ ID NO ("SEQ ID") assigned to each sequence for use in the present specification; 2) theCluster Identification No. ("CLUSTER"); 3) the Sequence Name assigned to each sequence; 3) the sequence name ("SEQ NAME") used as an internal identifier of the sequence; 4) the name assigned to the clone from which the sequence was isolated ("CLONE ID"); and 5) the name of the library from which the sequence was isolated ("LIBRARY"). Because at least some of the provided polynucleotides represent partial mRNA transcripts, two or more polynucleotides may represent different regions of the same mRNA transcript and the same gene and/or may be contained within the same clone. Thus, for example, if two or more SEQ ID NOS: are identified as belonging to the same clone, then either sequence can be used to obtain the full-length mRNA or gene. Clones which comprise the sequences described herein were deposited as set out in the tables indicated below (see Example entitled "Deposit Information").

Example 2: Contig Assembly

The sequences of the polynucleotides provided in the present invention can be used to extend the sequence information of the gene to which the polynucleotides correspond (*e.g.*, a gene, or mRNA encoded by the gene, having a sequence of the polynucleotide described herein). This expanded sequence information can in turn be used to further characterize the corresponding gene, which in turn provides additional information about the nature of the gene product (*e.g.*, the normal function of the gene product). The additional information can serve to provide additional evidence of the gene product's use as a therapeutic target, and provide further guidance as to the types of agents that can modulate its activity.

For example, a contig was assembled using the sequence of a polynucleotide described herein. A "contig" is a contiguous sequence of nucleotides that is assembled from nucleic acid sequences having overlapping (*e.g.*, shared or substantially similar) sequence information. The sequences of publicly-available ESTs (Expressed Sequence Tags) and the sequences of various of the above-described polynucleotides were used in the contig assembly. The contig was assembled using the software program Sequencher, version 4.05, according to the manufacturer's instructions. The sequence information obtained in the contig assembly was then used to obtain a consensus sequence derived from the contig using the Sequencher program. The resulting consensus sequence was used to search both the public databases as well as databases internal to the applicants to match the consensus polynucleotide with homology data and/or differential gene expressed data.

The final result provided the sequences listed as SEQ ID NOS: 1220-1428 in the accompanying Sequence Listing and summarized in Tables 3 and 4 (inserted prior to claims). Table 3 provides a summary of the consensus sequences assembled as described. Specifically, Table 3 provides: 1) the SEQ ID NO ("SEQ ID") assigned to each consensus sequence for use in the present specification; 2) theCluster Identification No. ("CLUSTER"); and 3) the consensus sequence name ("CONSENSUS SEQ NAME") used as an internal identifier of the sequence.

A correlation between the polynucleotide used in consensus sequence assembly as described above and the corresponding consensus sequence is contained in Table 4. Specifically Table 4 provides: 1) the SEQ ID NO of the consensus sequence ("CONSENSUS SEQ ID"); 2) the consensus sequence name ("CONSENSUS SEQ NAME") used as an internal identifier of the sequence; 3) the SEQ ID NO of the polynucleotide ("POLYNTD SEQ ID") of SEQ ID NOS: 1-1219 used in assembly of the consensus sequence; and 4) the sequence name ("POLYNTD SEQ NAME") of the polynucleotide of SEQ ID NOS: 1-1219 used in assembly of the consensus sequence.

Example 3: Additional Gene Characterization

Sequences of the polynucleotides of SEQ ID NOS: 1-1219 were used as a query sequence in a TeraBLASTN search of the DoubleTwist Human Genome Sequence Database (DoubleTwist, Inc., Oakland, CA), which contains all the human genomic sequences that have been assembled into a contiguous model of the human genome. Predicted cDNA and protein sequences were obtained where a polynucleotide of the invention was homologous to a predicted full-length gene sequence. Alternatively, a sequence of a contig or consensus sequence described herein could be used directly as a query sequence in a TeraBLASTN search of the DoubleTwist Human Genome Sequence Database.

The final results of the search provided the predicted cDNA sequences listed as SEQ ID NOS: 1429-1485 in the accompanying Sequence Listing and summarized in Table 5 (inserted prior to claims), and the predicted protein sequences listed as SEQ ID NOS:1486-1542 in the accompanying

Sequence Listing and summarized in Table 6 (inserted prior to claims). Specifically, Table 5 provides: 1) the SEQ ID NO ("SEQ ID") assigned to each cDNA sequence for use in the present specification; 2) the cDNA sequence name ("cDNA SEQ NAME") used as an internal identifier of the sequence; 3) the chromosome ("CHROM") containing the gene corresponding to the cDNA sequence; and 4) the exon ("EXON") of the gene corresponding to the cDNA sequence to which the polynucleotide of SEQ ID NOS: 1-1219 maps. Table 6 provides: 1) the SEQ ID NO ("SEQ ID") assigned to each protein sequence for use in the present specification; 2) the protein sequence name ("PROTEIN SEQ NAME") used as an internal identifier of the sequence; 3) the chromosome ("CHROM") containing the gene corresponding to the cDNA sequence; and 4) the exon ("EXON") of the gene corresponding to the cDNA and protein sequence to which the polynucleotide of SEQ ID NOS: 1-1219 maps.

A correlation between the polynucleotide used as a query sequence as described above and the corresponding predicted cDNA and protein sequences is contained in Table 7. Specifically Table 7 provides: 1) the SEQ ID NO of the cDNA ("cDNA SEQ ID"); 2) the cDNA sequence name ("cDNA SEQ NAME") used as an internal identifier of the sequence; 3) the SEQ ID NO of the protein ("PROTEIN SEQ ID") encoded by the cDNA sequence 4) the sequence name of the protein ("PROTEIN SEQ NAME") encoded by the cDNA sequence; 5) the SEQ ID NO of the polynucleotide ("POLYNTD SEQ ID") of SEQ ID NOS: 1-1219 that maps to the cDNA and protein; and 6) the sequence name ("POLYNTD SEQ NAME") of the polynucleotide of SEQ ID NOS: 1-1219 that maps to the cDNA and protein.

Through contig and consensus sequence assembly and the use of homology searching software programs, the sequence information provided herein can be readily extended to confirm, or confirm a predicted, gene having the sequence of the polynucleotides described in the present invention. Further the information obtained can be used to identify the function of the gene product of the gene corresponding to the polynucleotides described herein. While not necessary to the practice of the invention, identification of the function of the corresponding gene, can provide guidance in the design of therapeutics that target the gene to modulate its activity and modulate the cancerous phenotype (*e.g.*, inhibit metastasis, proliferation, and the like).

<u>Example 4:Results of Public Database Search to Identify Function of Gene Products</u>

SEQ ID NOS:1-1485 were translated in all three reading frames, and the nucleotide sequences and translated amino acid sequences used as query sequences to search for homologous sequences in the GenBank (nucleotide sequences) database. Query and individual sequences were aligned using the TeraBLAST program available from TimeLogic, Crystal Bay, Nevada. The sequences were masked to various extents to prevent searching of repetitive sequences or poly-A sequences, using the RepeatMasker masking program for masking low complexity as described above.

Table 8 (inserted prior to claims) provides the alignment summaries having a p value of 1 x 10e-2 or less indicating substantial homology between the sequences of the present invention and those of the indicated public databases. Specifically, Table 8 provides: 1) the SEQ ID NO ("SEQ ID") of the query sequence; 2) the sequence name ("SEQ NAME") used as an internal identifier of the query sequence; 3) the accession number ("ACCESSION") of the GenBank database entry of the homologous sequence; 4) a description of the GenBank sequences ("GENBANK DESCRIPTION"); and 5) the score of the similarity of the polynucleotide sequence and the GenBank sequence ("GENBANK SCORE"). The alignments provided in Table 8 are the best available alignment to a DNA sequence at a time just prior to filing of the present specification. Incorporated by reference is all publicly available information regarding the sequence listed in Table 8 and their related sequences. The search program and database used for the alignment, as well as the calculation of the p value are also indicated. Full length sequences or fragments of the polynucleotide sequences can be used as probes and primers to identify and isolate the full length sequence of the corresponding polynucleotide.

Example 5:Members of Protein Families

SEQ ID NOS:1-1219 were used to conduct a profile search as described in the specification above. Several of the polynucleotides of the invention were found to encode polypeptides having characteristics of a polypeptide belonging to a known protein family (and thus represent members of these protein families) and/or comprising a known functional domain. Table 9 (inserted prior to claims) provides: 1) the SEQ ID NO ("SEQ ID") of the query polynucleotide sequence; 2) the sequence name ("SEQ NAME") used as an internal identifier of the query sequence; 3) the name ("PFAM NAME") of the profile hit; 4) a brief description of the profile hit ("PFAM DESCRIPTION"); 5) the score ("SCORE") of the profile hit; 6) the starting nucleotide of the profile hit ("START"); and 7) the ending nucleotide of the profile hit ("END").

In addition, SEQ ID NOS:1486-1542 were also used to conduct a profile search as described above. Several of the polypeptides of the invention were found to have characteristics of a polypeptide belonging to a known protein family (and thus represent members of these protein families) and/or comprising a known functional domain. Table 10 (inserted prior to claims) provides: 1) the SEQ ID NO ("SEQ ID") of the query protein sequence; 2) the sequence name ("PROTEIN SEQ NAME") used as an internal identifier of the query sequence; 3) the name ("PFAM NAME") of the profile hit; 4) a brief description of the profile hit ("PFAM DESCRIPTION"); 5) the score ("SCORE") of the profile hit; 6) the starting residue of the profile hit ("START"); and 7) the ending residue of the profile hit ("END").

Some SEQ ID NOS exhibited multiple profile hits where the query sequence contains overlapping profile regions, and/or where the sequence contains two different functional domains. Each of the profile hits of Tables 9 and 10 is described in more detail below. The acronyms for the profiles (provided in parentheses) are those used to identify the profile in the Pfam, Prosite, and InterPro databases. The Pfam database can be accessed through web sites supported by Genome Sequencing Center at the Washington University School of Medicine or by the European Molecular Biology Laboratories in Heidelberg, Germany. The Prosite database can be accessed at the ExPASy Molecular Biology Server on the internet. The InterPro database can be accessed at a web site supported by the EMBL European Bioinformatics Institute. The public information available on the Pfam, Prosite, and InterPro databases regarding the various profiles, including but not limited to the activities, function, and consensus sequences of various proteins families and protein domains, is incorporated herein by reference.

Epidermal Growth Factor (EGF; Pfam Accession No. PF00008). SEQ ID NOS:417 and 418 represent polynucleotides encoding a member of the EGF family of proteins. The distinguishing characteristic of this family is the presence of a sequence of about thirty to forty amino acid residues found in epidermal growth factor (EGF) which has been shown to be present, in a more or less conserved form, in a large number of other proteins (Davis, *New Biol.* (1990) *2*:410-419; Blomquist *et al., Proc. Natl. Acad. Sci. U.S.A.* (1984) *81*:7363-7367; Barkert *et al., Protein Nucl. Acid Enz.* (1986) *29*:54-86; Doolittle *et al., Nature.* (1984) *307*:558-560; Appella *et al., FEBS Lett.* (1988) *231*:1-4; Campbell and Bork, *Curr. Opin. Struct. Biol.* (1993) *3*:385-392). A common feature of the domain is that the conserved pattern is generally found in the extracellular domain of membrane-bound proteins or in proteins known to be secreted. The EGF domain includes six cysteine residues which have been shown to be involved in disulfide bonds. The main structure is a two-stranded beta-sheet followed by a loop to a C-terminal short two-stranded sheet. Subdomains between the conserved cysteines strongly vary in length. These consensus patterns are used to identify members of this family: C-x-C-x(5)-G-x(2)-C and C-x-C-x(s)-[GP]-[FYW]-x(4,8)-C.

Seven Transmembrane Integral Membrane Proteins -- Rhodopsin Family (7tm_1; Pfam Accession No. PF00001). SEQ ID NO:321 corresponds to a sequence encoding a polypeptide that is a member of the seven transmembrane (7tm) receptor rhodopsin family. G-protein coupled receptors of the (7tm) rhodopsin family (also called R7G) are an extensive group of hormones, neurotransmitters, and light receptors which transduce extracellular signals by interaction with guanine nucleotide-binding (G) proteins (Strosberg, *Eur. J. Biochem.* (1991) 196:1; Kerlavage, *Curr. Opin. Struct. Biol.* (1991) 1:394; Probst et al., *DNA Cell Biol.* (1992) 11:1; Savarese et al., *Biochem. J.* (1992) 283:1. The consensus pattern that contains the conserved triplet and that also spans the major part of the third transmembrane helix is used to detect this widespread family of proteins: [GSTALIVMFYWC]-[GSTANCPDE]-{EDPKRH}-x(2)-[LIVMNQGA]-x(2)- [LIVMFT]-[GSTANC]-[LIVMFYWSTAC]-[DENH]-R-[FYWCSH]-x(2)- [LIVM].

Basic Region Plus Leucine Zipper Transcription Factors (bZIP; Pfam Accession No. PF00170). SEQ ID NO:638 represents a polynucleotide encoding a novel member of the family of basic region plus leucine zipper transcription factors. The bZIP superfamily (Hurst, *Protein Prof.* (1995) *2*:105; and Ellenberger, *Curr. Opin. Struct. Biol.* (1994) *4*:12) of eukaryotic DNA-binding transcription factors encompasses proteins that contain a basic region mediating sequence-specific DNA-binding followed by a leucine zipper required for dimerization. The consensus pattern for this protein family is: [KR]-x(1,3)-[RKSAQ]-N-x(2)-[SAQ](2)-x-[RKTAENQ]-x-R-x-[RK].

<u>Reverse Transcriptase (rvt; Pfam Accession No. PF00078).</u> SEQ ID NO:137 represents a polynucleotide encoding a reverse transcriptase, which occurs in a variety of mobile elements, including retrotransposons, retroviruses, group II introns, bacterial msDNAs, hepadnaviruses, and caulimoviruses (Xiong and Eickbush, *EMBO J* (1990) *9*:3353-3362). Reverse transcriptases catalyze RNA-template-directed extension of the 3'-end of a DNA strand by one deoxynucleotide at a time and require an RNA or DNA primer.

<u>KRAB box (KRAB; Pfam Accession No. PF01352).</u> SEQ ID NO:1012 represents a polypeptide having a Krueppel-associated box (KRAB). A KRAB box is a domain of around 75 amino acids that is found in the N-terminal part of about one third of eukaryotic Krueppel-type C2H2 zinc finger proteins (ZFPs). It is enriched in charged amino acids and can be divided into subregions A and B, which are predicted to fold into two amphipathic alpha-helices. The KRAB A and B boxes can be separated by variable spacer segments and many KRAB proteins contain only the A box.

The KRAB domain functions as a transcriptional repressor when tethered to the template DNA by a DNA-binding domain. A sequence of 45 amino acids in the KRAB A subdomain has been shown to be necessary and sufficient for transcriptional repression. The B box does not repress by itself but does potentiate the repression exerted by the KRAB A subdomain. Gene silencing requires the binding of the KRAB domain to the RING-B box-coiled coil (RBCC) domain of the KAP-1/TIF1-beta corepressor. As KAP-1 binds to the heterochromatin proteins HP1, it has been proposed that the KRAB-ZFP-bound target gene could be silenced following recruitment to heterochromatin.

KRAB-ZFPs constitute one of the single largest class of transcription factors within the human genome, and appear to play important roles during cell differentiation and development. The KRAB domain is generally encoded by two exons. The regions coded by the two exons are known as KRAB-A and KRAB-B.

<u>Armadillo/beta-catenin-like repeat (Armadillo_seg; Pfam Accession No. PF00514).</u> SEQ ID NO: 1486 represents a polypeptide having sequence similarity with the armadillo/beta-catenin-like repeat (armadillo). The armadillo repeat is an approximately 40 amino acid long tandemly repeated sequence motif first identified in the Drosophila segment polarity gene armadillo. Similar repeats were later found in the mammalian armadillo homolog beta-catenin, the junctional plaque protein plakoglobin, the adenomatous polyposis coli (APC) tumor suppressor protein, and a number of other proteins (Peifer *et al., Cell* 76(2):786-791 (1994)).

The 3 dimensional fold of an armadillo repeat is known from the crystal structure of beta-catenin (Rojas *et al., Cell* 95:105-130 (1998)). There, the 12 repeats form a superhelix of alpha-helices, with three helices per unit. The cylindrical structure features a positively charged grove which presumably interacts with the acidic surfaces of the known interaction partners of beta-catenin.

Cadherin domain (cadherin; Pfam Accession No. PF00028). SEQ ID NO: 1523 represents a polypeptide having sequence similarity to a cadherin domain. Cadherins are a family of animal glycoproteins responsible for calcium-dependent cell-cell adhesion (Takeichi, *Annu. Rev. Biochem.* 59:237-252(1990); Takeichi, *Trends Genet.* 3:213-217(1987)). Cadherins preferentially interact with themselves in a homophilic manner in connecting cells; thus acting as both receptor and ligand. A wide number of tissue-specific forms of cadherins are known, for example: Epithelial (E-cadherin) (CDH1); Neural (N-cadherin) (CDH2); Placental (P-cadherin) (CDH3); Retinal (R-cadherin) (CDH4); Vascular endothelial (VE-cadherin) (CDH5); Kidney (K-cadherin) (CDH6); Cadherin-8 (CDH8); Cadherin-9 (CDH9); Osteoblast (OB-cadherin) (CDH11); Brain (BR-cadherin) (CDH12); T-cadherin (truncated cadherin) (CDH13); Muscle (M-cadherin) (CDH15); Kidney (Ksp-cadherin) (CDH16); and Liver-intestine (LI-cadherin) (CDH17).

Structurally, cadherins are built of the following domains: a signal sequence, followed by a propeptide of about 130 residues, then an extracellular domain of around 600 residues, then a transmembrane region, and finally a C-terminal cytoplasmic domain of about 150 residues. The extracellular domain can be sub-divided into five parts: there are four repeats of about 110 residues followed by a region that contains four conserved cysteines. The calcium-binding region of cadherins may be located in the extracellular repeats. The signature pattern for the repeated domain is located in the C-terminal extremity, which is its best conserved region. The pattern includes two conserved aspartic acid residues and two asparagines; these residues could be implicated in the binding of calcium. The consensus pattern is: [LIV]-x-[LIV]-x-D-x-N-D-[NH]-x-P.

CBS domain (CBS; Pfam Accession No. PF00571). SEQ ID NOS:1510 and 1511 represent polypeptides having sequence similarity to CBS domains, which are present in all 3 forms of cellular life, including two copies in inosine monophosphate dehydrogenase, of which one is disordered in the crystal structure. A number of disease states are associated with CBS-containing proteins including homocystinuria, Becker's and Thomsen disease.

CBS domains are small intracellular modules of unknown function. They are mostly found in 2 or four copies within a protein. Pairs of CBS domains dimerise to form a stable globular domain (Zhang et al., *Biochemistry* 38:4691-4700 (1999)). Two CBS domains are found in inosine-monophosphate dehydrogenase from all species, however the CBS domains are not needed for activity. CBS domains are found attached to a wide range of other protein domains suggesting that CBS domains may play a regulatory role. The region containing the CBS domains in Cystathionine-beta synthase is involved in regulation by S-AdoMet (Zhang et al., *Biochemistry* 38:4691-4700

(1999)). The 3D Structure is found as a sub-domain in TIM barrel of inosine-monophosphate dehydrogenase.

Phorbol esters/diacylglycerol binding domain (C1 domain) (DAG_PE-bind; Pfam Accessin No. PF00130). SEQ ID NO: 1514 represents a polypeptide having sequence similarity to the Phorbol esters/diacylglycerol binding domain (C1 domain). Diacylglycerol (DAG) is an important second messenger. Phorbol esters (PE) are analogues of DAG and potent tumor promoters that cause a variety of physiological changes when administered to both cells and tissues. DAG activates a family of serine/threonine protein kinases, collectively known as protein kinase C (PKC) (Azzi *et al., Eur. J. Biochem.* 208:547-557 (1992)). Phorbol esters can also directly stimulate PKC.

The N-terminal region of PKC, known as C1, has been shown to bind PE and DAG in a phospholipid and zinc-dependent fashion(Ono *et al., Proc. Natl. Acad. Sci. U.S.A.* 86:4868-4871 (1989)). The C1 region contains one or two copies (depending on the isozyme of PKC) of a cysteine-rich domain about 50 amino-acid residues long and essential for DAG/PE-binding. The DAG/PE-binding domain binds two zinc ions; the ligands of these metal ions are probably the six cysteines and two histidines that are conserved in the C1 domain. The consensus sequence for the C1 domain is: H-x-[LIVMFYW]-x(8,11)-C-x(2)-C-x(3)-[LIVMFC]-x(5,10)-C-x(2)-C-x(4)-[HD]-x(2)-C-x(5,9)-C [All the C and H are involved in binding Zinc].

GATA zinc finger (GATA; Pfam Accession No. PF00320). SEQ ID NO:1520 represents a polypeptide having sequence similarity to GATA zinc finger. A number of transcription factors, including erythroid-specific transcription factor and nitrogen regulatory proteins, specifically bind the DNA sequence (A/T)GATA(A/G) in the regulatory regions of genes (Yamamoto *et al., Genes Dev.* 4:1650-1662 (1990)) and are consequently termed GATA-binding transcription factors. The interactions occur via highly-conserved zinc finger domains in which the zinc ion is coordinated by 4 cysteine residues (Evans and Felsenfeld, *Cell* 58:877-885 (1989); Omichinski et al., *Science* 261:438-446 (1993)).

NMR studies have shown the core of the zinc finger to comprise 2 irregular anti-parallel beta-sheets and an alpha-helix, followed by a long loop to the C-terminal end of the finger. The N-terminal part, which includes the helix, is similar in structure, but not sequence, to the N-terminal zinc module of the glucocorticoid receptor DNA-binding domain. The helix and the loop connecting the 2 beta-sheets interact with the major groove of the DNA, while the C-terminal tail wraps around into the minor groove. It is this tail that is the essential determinant of specific binding. Interactions between the zinc finger and DNA are mainly hydrophobic, explaining the preponderance of thymines in the binding site; a large number of interactions with the phosphate backbone have also been observed (Omichinski et al., *Science* 261:438-446 (1993)). Two GATA zinc fingers are found in the GATA transcription factors; however, there are several proteins which only contains a single copy of the domain. The consensus sequence of the domain is: C-x-[DN]-C-x(4,5)-[ST]-x(2)-W-[HR]-[RK]-x(3)-[GN]-x(3,4)- C-N-[AS]-C [The four C's are zinc ligands].

Glutathione S-transferase, N-terminal domain (GST_N; Pfam Accession No. PF02798). SEQ ID NO: 1507 represents a polypeptide having sequence similarity to Glutathione S-transferase, N-terminal domain. In eukaryotes, glutathione S-transferases (GSTs) participate in the detoxification of reactive electrophilic compounds by catalysing their conjugation to glutathione. The GST domain is also found in S-crystallins from squid, and proteins with no known GST activity, such as eukaryotic elongation factors 1-gamma and the HSP26 family of stress-related proteins, which include auxin-regulated proteins in plants and stringent starvation proteins in E. coli. The major lens polypeptide of Cephalopoda is also a GST.

Bacterial GSTs of known function often have a specific, growth-supporting role in biodegradative metabolism: epoxide ring opening and tetrachlorohydroquinone reductive dehalogenation are two examples of the reactions catalysed by these bacterial GSTs. Some regulatory proteins, like the stringent starvation proteins, also belong to the GST family. GST seems to be absent from Archaea in which gamma-glutamylcysteine substitute to glutathione as major thiol.

Glutathione S-transferases form homodimers, but in eukaryotes can also form heterodimers of the A1 and A2 or YC1 and YC2 subunits. The homodimeric enzymes display a conserved structural fold. Each monomer is composed of a distinct N-terminal sub-domain, which adopts the thioredoxin fold, and a C-terminal all-helical sub-domain.

GTF2I-like repeat (GTF2I; Pfam Accession No. PF02946). SEQ ID NOS:1500, 1501, and 1542 represent polypeptides having sequence similarity to proteins having GTF2I-like repeat. This region of sequence similarity is found up to six times in a variety of proteins including GTF2I. It has been suggested that this may be a DNA binding domain (O'Mahoney *et al., Mol. Cell. Biol.* 18:6641-6652 (1998); Osborne *et al., Genomics* 57:279-284 (1999)).

Core histone H2A/H2B/H3/H4 (histone; Pfam Accession No. PF00125). SEQ ID NO:1497 represents a polypeptide having sequence similarity to core histone H2A/H2B/H3/H4 family polypeptides. Histone H2A is one of the four histones, along with H2B, H3 and H4, which forms the eukaryotic nucleosome core. Using alignments of histone H2A sequences (Wells and Brown, *Nucleic Acids Res.* 19:2173-2188(1991); Thatcher and Gorovsky, *Nucleic Acids Res.* 22:174-179(1994)) a conserved region in the N-terminal part of H2A was used to develop a signature pattern. This region is conserved both in classical S-phase regulated H2A's and in variant histone H2A's which are synthesized throughout the cell cycle. The consensus pattern is: [AC]-G-L-x-F-P-V.

Histone H4, along with H3, plays a central role in nucleosome formation. The sequence of histone H4 has remained almost invariant in more then 2 billion years of evolution (Thatcher and Gorovsky, *Nucleic Acids Res.* 22:174-179(1994)). The region used as a signature pattern is a pentapeptide found in positions 14 to 18 of all H4 sequences. It contains a lysine residue which is often acetylated (Doenecke and Gallwitz, *Mol. Cell. Biochem.* 44:113-128(1982)) and a histidine residue which is implicated in DNA-binding (Ebralidse *et al., Nature* 331:365-367(1988)). The consensus pattern is: G-A-K-R-H.

Histone H3 is a highly conserved protein of 135 amino acid residues (Wells and Brown, *Nucleic Acids Res.* 19:2173-2188(1991); Thatcher and Gorovsky, *Nucleic Acids Res.* 22:174-179(1994)). Two signature patterns have been developed, the first one corresponds to a perfectly conserved heptapeptide in the N-terminal part of H3, while the second one is derived from a conserved region in the central section of H3. The consensus patterns are: K-A-P-R-K-Q-L and P-F-x-[RA]-L-[VA]-[KRQ]-[DEG]-[IV].

The signature pattern of histone H2B corresponds to a conserved region in the C-terminal part of the protein. The consensus pattern is: [KR]-E-[LIVM]-[EQ]-T-x(2)-[KR]-x-[LIVM](2)-x-[PAG]-[DE]-L-x-[KR]-H-A-[LIVM]-[STA]-E-G

<u>HMG (high mobility group) box (HMG_box; Pfam Accession No. PF00505).</u> SEQ ID NO:1525 corresponds to a polypeptide having sequence similarity to high mobility group proteins, a family of relatively low molecular weight non-histone components in chromatin. HMG1 (also called HMG-T in fish) and HMG2 (Bustin *et al., Biochim. Biophys. Acta* 1049: 231-243(1990)) are two highly related proteins that bind single-stranded DNA preferentially and unwind double-stranded DNA. HMG1/2 have about 200 amino acid residues with a highly acidic C-terminal section which is composed of an uninterrupted stretch of from 20 to 30 aspartic and glutamic acid residues; the rest of the protein sequence is very basic. In addition to the HMG1 and HMG2 proteins, HMG-domains occur in single or multiple copies in the following protein classes; the SOX family of transcription factors; SRY sex determining region Y protein and related proteins; LEF1 lymphoid enhancer binding factor 1; SSRP recombination signal recognition protein; MTF1 mitochondrial transcription factor 1; UBF1/2 nucleolar transcription factors; Abf2 yeast ARS-binding factor; and yeast transcription factors Ixr1, Rox1, Nhp6a, Nhp6b and Spp41.

<u>Importin beta binding domain (IBB; Pfam Accession No. PF01749).</u> SEQ ID NO: 1486 represents a polypeptide having sequence similarity to importin beta binding domain family polypeptides. This family consists of the importin alpha (karyopherin alpha), importin beta (karyopherin beta) binding domain. The domain mediates formation of the importin alpha beta complex; required for classical NLS import of proteins into the nucleus, through the nuclear pore complex and across the nuclear envelope. Also in the alignment is the NLS of importin alpha which overlaps with the IBB domain (Moroianu *et al., Proc. Natl. Acad. Sci. U.S.A.* 93:6572-6576(1996)).

<u>T-box domain (T-box; Pfam Accession No. PF00907).</u> SEQ ID NOS:1518 represents a polypeptide having sequence similarity to proteins having a T-box domain. The T-box gene family is an ancient group of putative transcription factors that appear to play a critical role in the development of all animal species. These genes were uncovered on the basis of similarity to the DNA binding

domain (Papaioannou and Silver, *Bioessays* 20:9-19 (1998)) of murine Brachyury (T) gene product, which similarity is the defining feature of the family. The Brachyury gene is named for its phenotype, which was identified 70 years ago as a mutant mouse strain with a short blunted tail. The gene, and its paralogues, have become a well-studied model for the family, and hence much of what is known about the T-box family is derived from the murine Brachyury gene.

Consistent with its nuclear location, Brachyury protein has a sequence-specific DNA-binding activity and can act as a transcriptional regulator (Wattler *et al., Genomics* 48:24-33(1998)). Homozygous mutants for the gene undergo extensive developmental anomalies, thus rendering the mutation lethal (Kavka and Green, *Biochim. Biophys. Acta* 1333(2) (1997)). The postulated role of Brachyury is as a transcription factor, regulating the specification and differentiation of posterior mesoderm during gastrulation in a dose-dependent manner (Papaioannou and Silver, *Bioessays* 20:9-19 (1998)).

Common features shared by T-box family members are, DNA-binding and transcriptional regulatory activity, a role in development and conserved expression patterns. Most of the known genes in all species are expressed in mesoderm or mesoderm precursors (Papaioannou, *Trends Genet.* 13:212-213(1997)). Members of the T-box family contain a domain of about 170 to 190 amino acids known as the T-box domain (Papaioannou, *Trends Genet.* 13: 212-213(1997); Bollag *et al., Nat. Genet.* 7: 383-389(1994); Agulnik et al., *Genetics* 144:249-254(1996)) and which probably binds DNA. As signature patterns for the T-domain, we selected two conserved regions. The first region corresponds to the N-terminal of the domain and the second one tothe central part. The consensus sequences are: L-W-x(2)-[FC]-x(3,4)-[NT]-E-M-[LIV](2)-T-x(2)-G-[RG]-[KRQ] and [LIVMFYW]-H-[PADH]-[DENQ]-[GS]-x(3)-G-x(2)-W-M-x(3)-[IVA]-x-F.

<u>60s Acidic ribosomal protein (60s_ribosomal; Pfam Accession No. PF00428).</u> SEQ ID NO: 905 represents a polynucleotide encoding a member of the 60s acidic ribosomal protein family. The 60S acidic ribosomal protein plays an important role in the elongation step of protein synthesis. This family includes archaebacterial L12, eukaryotic P0, P1 and P2 (Remacha *et al., Biochem. Cell Biol.* 73:959-968(1995)).

Some of the proteins in this family are allergens. A nomenclature system has been established for antigens (allergens) that cause IgE-mediated atopic allergies in humans (WHO/IUIS Allergen Nomenclature Subcommittee King T.P., Hoffmann D., Loewenstein H., Marsh D.G., Platts-Mills T.A.E., Thomas W. Bull. World Health Organ. 72:797-806(1994)). This nomenclature system is defined by a designation that is composed of the first three letters of the genus; a space; the first letter of the species name; a space and an arabic number. In the event that two species names have identical designations, they are discriminated from one another by adding one or more letters (as necessary) to each species designation. The allergens in this family include allergens with the following designations: Alt a 6, Alt a 12, Cla h 3, Cla h 4, and Cla h 12.

AP endonuclease family 1 (AP_endonucleas1; Pfam Accession No. PF01260). SEQ ID NOS:358 and 836 correspond to a polynucleotide encoding a member of the family of polypeptides designated AP endonuclease family 1. DNA damaging agents such as the antitumor drugs bleomycin and neocarzinostatin or those that generate oxygen radicals produce a variety of lesions in DNA. Amongst these is base-loss which forms apurinic/apyrimidinic (AP) sites or strand breaks with atypical 3'-termini. DNA repair at the AP sites is initiated by specific endonuclease cleavage of the phosphodiester backbone. Such endonucleases are also generally capable of removing blocking groups from the 3'-terminus of DNA strand breaks.

AP endonucleases can be classified into two families on the basis of sequence similarity. This family contains members of AP endonuclease family 1. Except for Rrp1 and arp, these enzymes are proteins of about 300 amino-acid residues. Rrp1 and arp both contain additional and unrelated sequences in their N-terminal section (about 400 residues for Rrp1 and 270 for arp). The proteins contain glutamate which has been shown (Mol *et al., Nature* 374: 381-386(1995)), in the Escherichia coli enzyme to bind a divalent metal ion such as magnesium or manganese. The consensus sequences for this family of polypeptides are: [APF]-D-[LIVMF](2)-x-[LIVM]-Q-E-x-K [E binds a divalent metal ion]; D-[ST]-[FY]-R-[KH]-x(7,8)-[FYW]-[ST]-[FYW](2); and N-x-G-x-R-[LIVM]-D-[LIVMFYH]-x-[LV]-x-S

Bowman-Birk serine protease inhibitor family (Bowman-Birk_leg; Pfam Accession No. 00228). SEQ ID NO: 321 represents a polynucleotide encoding a polypeptide having sequence similarity to a member of the Bowman-Birk serine protease inhibitor family. The Bowman-Birk inhibitor family (Laskowski and Kato, *Annu. Rev. Biochem.* 49:593-626(1980)) is one of the numerous families of serine proteinase inhibitors and has a duplicated structure and generally possesses two distinct inhibitory sites.

These inhibitors are found in the seeds of all leguminous plants as well as in cereal grains. In cereals they exist in two forms, one of which is a duplication of the basic structure (Tashiro *et al., J. Biochem.* 102:297-306(1987)). The signature pattern for sequences belonging to this family of inhibitors is in the central part of the domain and includes four cysteines. The consensus pattern is: C-x(5,6)-[DENQKRHSTA]-C-[PASTDH]-[PASTDK]-[ASTDV]-C-[NDEKS]-[DEKRHSTA]-C [The four C's are involved in disulfide bonds]. Note that this pattern can be found twice in some duplicated cereal inhibitors.

Cation efflux family (Cation_efflux; Pfam Accession No. PF01545). SEQ ID NO: 321 encodes a polypeptide having sequence similarity to members of the cation efflux family of proteins. Members of this family are integral membrane proteins, that are found to increase tolerance to divalent metal ions such as cadmium, zinc, and cobalt. These proteins are thought to be efflux pumps that remove these ions from cells (Xiong and Jayaswal, *J. Bacteriol.* 180: 4024-4029(1998); Kunito *et al, Biosci. Biotechnol. Biochem.* 60: 699-704(1996)).

DC1 domain (DC1; Pfam Accession No. PF03107). SEQ ID NO: 89 corresponds to a polypeptide having sequence similarity to a DC1 domain. This short domain is rich in cysteines and histidines. The pattern of conservation is similar to that found in DAG_PE-bind (Pfam Accession No. PF00130), therefore this domain has been termed DC1 for divergent C1 domain. Like the DAG_PE-bind domain, this domain probably also binds to two zinc ions. The function of proteins with this domain is uncertain, however this domain may bind to molecules such as diacylglycerol. This family are found in plant proteins.

Pneumovirus attachment glycoprotein G (Glycoprotein_G; Pfam Accession No. PF00802). SEQ ID NO:995 represents a polypeptide having sequence similarity to members of the Pneumovirus attachment glycoprotein G protein family. This family includes attachment proteins from respiratory synctial virus. Glycoprotein G has not been shown to have any neuraminidase or hemagglutinin activity. The amino terminus is thought to be cytoplasmic, and the carboxyl terminus extracellular. The extracellular region contains four completely conserved cysteine residues.

NADH-Ubiquinone/plastoquinone (complex I), various chains (oxidored_q1; Pfam Accession No. PF00361). SEQ ID NO:413 represents a polypeptide having sequence similarity to NADH-Ubiquinone/plastoquinone (complex I), various chains protein family. This family is part of the NADH:ubiquinone oxidoreductase (complex I) which catalyses the transfer of two electrons from NADH to ubiquinone in a reaction that is associated with proton translocation across the membrane (Walker, *Q. Rev. Biophys.* 25: 253-324(1992)). Sub-families within this protein family include NADH-ubiquinone oxidoreductase chain 5; NADH-ubiquinone oxidoreductase chain 2; NADH-ubiquinone oxidoreductase chain 4; and Multicomponent K+:H+antiporter.

Protamine P1 (protamine_P1; Pfam Accession No. PF00260). SEQ ID NOS:645 and 1217 represent polypeptides having sequence similarity to Protamine P1 protein family. Protamines are small, highly basic proteins, that substitute for histones in sperm chromatin during the haploid phase of spermatogenesis. They pack sperm DNA into a highly condensed, stable and inactive complex. There are two different types of mammalian protamine, called P1 and P2. P1 has been found in all species studied, while P2 is sometimes absent. There also seems to be a single type of avian protamine whose sequence is closely related to that of mammalian P1 (Oliva *et al., J. Biol. Chem.* 264:17627-17630(1989)). A conserved region at the N-terminal extremity of the sequence is used as a signature pattern for this family of proteins. The consensus pattern is: [AV]-R-[NFY]-R-x(2,3)-[ST]-x-S-x-S.

Squash family serine protease inhibitor (squash; Pfam Accession No. PF00299). SEQ ID NO:995 represents a polypeptide having sequence similarity to Squash family serine protease inhibitor proteins. The squash inhibitors form one of a number of serine protease inhibitor families. The proteins, found in the seeds of cucurbitaceae plants (squash, cucumber, balsam pear, etc.), are approximately 30 residues in length, and contain 6 Cys residues, which form 3 disulfide bonds (Bode *et al., FEBS Lett.* 242: 285-292(1989)). The inhibitors function by being taken up by a serine protease (such as trypsin), which cleaves the peptide bond between Arg/Lys and Ile residues in the N-terminal portion of the protein (Bode *et al., FEBS Lett.* 242: 285-292(1989); Krishnamoorthi et al., Biochemistry 31: 898-904(1992)). Structural studies have shown that the inhibitor has an ellipsoidal shape, and is largely composed of beta-turns (Bode *et al., FEBS Lett.* 242: 285-292(1989)). The fold and Cys connectivity of the proteins resembles that of potato carboxypeptidase A inhibitor (Krishnamoorthi *et al., Biochemistry* 31: 898-904(1992)). The pattern used to detect this family of proteins spans the major part of the sequence and includes five of the six cysteines involved in disulfide bonds. The consensus pattern is: C-P-x(5)-C-x(2)-[DN]-x-D-C-x(3)-C-x-C [The five C's are involved in disulfide bonds]

Metallothionein family 5 (Metallothio_5; Pfam Accession No. PF02067). SEQ ID NO:995 represents a polypeptide having sequence similarity to metallothionein family 5 proteins. Metallothioneins (MT) are small proteins that bind heavy metals, such as zinc, copper, cadmium, and nickel. They have a high content of cysteine residues that bind the metal ions through clusters of thiolate bonds (Kagi, *Meth. Enzymol.* 205: 613-626(1991); Kagi and Kojima, *Experientia Suppl.* 52: 25-61(1987); Kagi and Schaffer, *Biochemistry* 27: 8509-8515(1988)).

Due to limitations in the original classification system of MTs, which did not allow clear differentiation of patterns of structural similarities, either between or within classes, all class I and class II MTs (the proteinaceous sequences) have now been grouped into families of phylogenetically-related and thus alignable sequences. Diptera (Drosophila, family 5) MTs are 40-43 residue proteins that contain 10 conserved cysteines arranged in five Cys-X-Cys groups. In particular, the consensus pattern C-G-x(2)-C-x-C-x(2)-Q-x(5)-C-x-C-x(2)-D-C-x-C has been found to be diagnostic of family 5 MTs. The protein is found primarily in the alimentary canal, and its induction is stimulated by ingestion of cadmium or copper (Lastowski *et al., J. Biol. Chem.* 260: 1527-1530(1985)). Mercury, silver and zinc induce the protein to a lesser extent.

Caenorhabditis. elegans Sre G protein-coupled chemoreceptor (Sre; Pfam Accession No. PF03125). SEQ ID NO:591 represents a polypeptide having sequence similarity to C. elegans Sre G protein-coupled chemoreceptor family proteins. C. elegans Sre proteins are candidate chemosensory receptors. There are four main recognized groups of such receptors: Odr-10, Sra, Sro, and Srg. Sre (this family), Sra Sra and Srb Srb comprise the Sra group. All of the above receptors are thought to be G protein-coupled seven transmembrane domain proteins (Troemel, *Bioessays* 21:1011-1020 (1999); Troemel *et al., Cell* 83:207-218 (1995)).

Syndecan domain (Syndecan; Pfam Accession No. PF01034). SEQ ID NO:995 corresponds to a polypeptide having a syndecan domain. Syndecans (Bernfield *et al., Annu. Rev. Cell Biol.* 8:365-393(1992); David, FASEB J. 7:1023-1030(1993)) are a family of transmembrane heparan sulfate proteoglycans which are implicated in the binding of extracellular matrix components and growth factors. Syndecans bind a variety of molecules via their heparan sulfate chains and can act as receptors or as co-receptors. Structurally, these proteins consist of four separate domains: a) a signal sequence; b) an extracellular domain (ectodomain) of variable length containing the sites of attachment of the heparan sulfate glycosaminoglycan side chains and whose sequence is not evolutionarily conserved in the various forms of syndecans; c) a transmembrane region; and d) a highly conserved cytoplasmic domain of about 30 to 35 residues which could interact with cytoskeletal proteins.

The signature pattern for syndecans starts with the last residue of the transmembrane region and includes the first 10 residues of the cytoplasmic domain. This region, which contains four basic residues, may act as a stop transfer site. The consensus pattern is: [FY]-R-[IM]-[KR]-K(2)-D-E-G-S-Y.

<u>L1 transposable element (Transposase_22; Pfam Accession No.PF02994).</u> SEQ ID NO:774 represents a polypeptide having an L1 transposable element. Many human L1 elements are capable of retrotransposition and some of these have been shown to exhibit reverse transcriptase (RT) activity (Sassaman *et al., Nat Genet* 16(1):37-43(1997)) although the function of many are, as yet, unknown. There are estimated to be 30-60 active L1 elements reside in the average diploid genome.

<u>WW domain (WW; Pfam Accession No. PF00397).</u> SEQ ID NO:431 represents a polypeptide having WW domain. The WW domain (also known as rsp5 or WWP) is a short conserved region in a number of unrelated proteins, among them dystrophin, responsible for Duchenne muscular dystrophy. This short domain may be repeated up to four times in some proteins (Bork and Sudol, *Trends Biochem. Sci.* 19: 531-533(1994); Andre and Springael, *Biochem. Biophys. Res. Commun.* 205: 1201-1205(1994); Hofmann and Bucher, *FEBS Lett.* 358: 153-157(1995); Sudol et al., *FEBS Lett.* 369: 67-71(1995)). The WW domain binds to proteins with particular proline-motifs, [AP]-P-P-[AP]-Y, and having four conserved aromatic positions that are generally Trp (Chen and Sudol, *Proc. Natl. Acad. Sci. U.S.A.* 92: 7819-7823(1995)). The name WW or WWP derives from the presence of these Trp as well as that of a conserved Pro. The WW domain is frequently associated with other domains typical for proteins in signal transduction processes.

A large variety of proteins containing the WW domain are known. These include; dystrophin, a multidomain cytoskeletal protein; utrophin, a dystrophin-like protein of unknown function; vertebrate YAP protein, substrate of an unknown serine kinase; mouse NEDD-4, involved in the embryonic development and differentiation of the central nervous system; yeast RSP5, similar to NEDD-4 in its molecular organization; rat FE65, a transcription-factor activator expressed preferentially in liver; tobacco DB10 protein and others. The consensus pattern is: W-x(9,11)-[VFY]-[FYW]-x(6,7)-[GSTNE]-[GSTQCR]-[FYW]-x(2)-P.

<u>Example 6: Detection of Differential Expression Using Arrays and source of patient tissue samples</u>

mRNA isolated from samples of cancerous and normal breast and colon tissue obtained from patients were analyzed to identify genes differentially expressed in cancerous and normal cells. Normal and cancerous tissues were collected from patients using laser capture microdissection (LCM) techniques, which techniques are well known in the art (see, *e.g.*, Ohyama *et al.* (2000) *Biotechniques* 29:530-6; Curran *et al.* (2000) *Mol. Pathol.* 53:64-8; Suarez-Quian *et al.* (1999) *Biotechniques* 26:328-35; Simone *et al.* (1998) *Trends Genet* 14:272-6; Conia *et al.* (1997) *J. Clin. Lab. Anal.* 11:28-38; Emmert-Buck *et al.* (1996) *Science* 274:998-1001).

Table 11 (inserted prior to claims) provides information about each patient from which colon tissue samples were isolated, including: the Patient ID ("PT ID") and Path ReportID ("Path ID"), which are numbers assigned to the patient and the pathology reports for identification purposes; the group ("Grp") to which the patients have been assigned; the anatomical location of the tumor ("Anatom Loc"); the primary tumor size ("Size"); the primary tumor grade ("Grade"); the identification of the histopathological grade ("Histo Grade"); a description of local sites to which the tumor had invaded ("Local Invasion"); the presence of lymph node metastases ("Lymph Met"); the incidence of lymph node metastases (provided as a number of lymph nodes positive for metastasis over the number of lymph nodes examined) ("Lymph Met Incid"); the regional lymphnode grade ("Reg Lymph Grade"); the identification or detection of metastases to sites distant to the tumor and their location ("Dist Met & Loc"); the grade of distant metastasis ("Dist Met Grade"); and general comments about the patient or the tumor ("Comments"). Histophatology of all primary tumors indicated the tumor was adenocarcinmoa except for Patient ID Nos. 130 (for which no information was provided), 392 (in which greater than 50% of the cells were mucinous carcinoma), and 784 (adenosquamous carcinoma). Extranodal extensions were described in three patients, Patient ID Nos. 784, 789, and 791. Lymphovascular invasion was described in Patient ID Nos. 128, 278, 517, 534, 784, 786, 789, 791, 890, and 892. Crohn's-like infiltrates were described in seven patients, Patient ID Nos. 52, 264, 268, 392, 393, 784, and 791.

Table 12 (below) provides information about each patient from which the breast tissue samples were isolated, including: 1) the "Pat Num", a number assigned to the patient for identification purposes; 2) the "Histology", which indicates whether the tumor was characterized as an intraductal carcinoma (IDC) or ductal carcinoma in situ (DCIS); 3) the incidence of lymph node metastases (LMF), represented as the number of lymph nodes positive to metastases out of the total number examined in the patient; 4) the "Tumor Size"; 5) "TNM Stage", which provides the tumor grade (T#), where the number indicates the grade and "p" indicates that the tumor grade is a pathological classification; regional lymph node metastasis (N#), where "0" indicates no lymph node metastases were found, "1" indicates lymph node metastases were found, and "X" means information not available and; the identification or detection of metastases to sites distant to the tumor and their location (M#), with "X" indicating that no distant mesatses were reported; and the stage of the tumor ("Stage Grouping"). "nr" indicates "no reported".

**Table 12.** Breast cancer patient data.

| Pat Num | Histology | LMF | Tumor Size | TNM Stage | Stage Grouping |
|---|---|---|---|---|---|
| 280 | IDC, DCIS+D2 | nr | 2 cm | T2NXMX | probable Stage II |
| 284 | IDC, DCIS | 0/16 | 2 cm | T2pN0MX | Stage II |
| 285 | IDC, DCIS | nr | 4.5 cm | T2NXMX | probable Stage II |
| 291 | IDC, DCIS | 0/24 | 4.5 cm | T2pN0MX | Stage II |
| 302 | IDC, DCIS | nr | 2.2 cm | T2NXMX | probable Stage II |
| 375 | IDC, DCIS | nr | 1.5 cm | T1NXMX | probable Stage I |
| 408 | IDC | 0/23 | 3.0 cm | T2pN0MX | Stage II |
| 416 | IDC | 0/6 | 3.3 cm | T2pN0MX | Stage II |
| 421 | IDC, DCIS | nr | 3.5 cm | T2NXMX | probable Stage II |
| 459 | IDC | 2/5 | 4.9 cm | T2pN1MX | Stage II |
| 465 | IDC | 0/10 | 6.5 cm | T3pN0MX | Stage II |
| 470 | IDC, DCIS | 0/6 | 2.5 cm | T2pN0MX | Stage II |
| 472 | IDC, DCIS | 6/45 | 5.0+ cm | T3pN1MX | Stage III |
| 474 | IDC | 0/18 | 6.0 cm | T3pN0MX | Stage II |
| 476 | IDC | 0/16 | 3.4 cm | T2pN0MX | Stage II |
| 605 | IDC, DCIS | 1/25 | 5.0 cm | T2pN1MX | Stage II |
| 649 | IDC, DCIS | 1/29 | 4.5 cm | T2pN1MX | Stage II |

<u>Identification of differentially expressed genes</u>

cDNA probes were prepared from total RNA isolated from the patient cells described above. Since LCM provides for the isolation of specific cell types to provide a substantially homogenous cell sample, this provided for a similarly pure RNA sample.

Total RNA was first reverse transcribed into cDNA using a primer containing a T7 RNA polymerase promoter, followed by second strand DNA synthesis. cDNA was then transcribed *in vitro* to produce antisense RNA using the T7 promoter-mediated expression (see, *e.g.*, Luo *et al.* (1999) *Nature Med* 5:117-122), and the antisense RNA was then converted into cDNA. The second set of cDNAs were again transcribed *in vitro*, using the T7 promoter, to provide antisense RNA. Optionally, the RNA was again converted into cDNA, allowing for up to a third round of T7-mediated amplification to produce more antisense RNA. Thus the procedure provided for two or three rounds of *in vitro* transcription to produce the final RNA used for fluorescent labeling.

Fluorescent probes were generated by first adding control RNA to the antisense RNA mix, and producing fluorescently labeled cDNA from the RNA starting material. Fluorescently labeled cDNAs prepared from the tumor RNA sample were compared to fluorescently labeled cDNAs prepared from normal cell RNA sample. For example, the cDNA probes from the normal cells were labeled with Cy3 fluorescent dye (green) and the cDNA probes prepared from the tumor cells were labeled with Cy5 fluorescent dye (red), and vice versa.

Each array used had an identical spatial layout and control spot set. Each microarray was divided into two areas, each area having an array with, on each half, twelve groupings of 32 x 12 spots, for a total of about 9,216 spots on each array. The two areas are spotted identically which provide for at least two duplicates of each clone per array.

Polynucleotides for use on the arrays were obtained from both publicly available sources and from cDNA libraries generated from selected cell lines and patient tissues. PCR products of from about 0.5kb to 2.0 kb amplified from these sources were spotted onto the array using a Molecular Dynamics Gen III spotter according to the manufacturer's recommendations. The first row of each of the 24 regions on the array had about 32 control spots, including 4 negative control spots and 8 test polynucleotides. The test polynucleotides were spiked into each sample before the labeling reaction with a range of concentrations from 2-600 pg/slide and ratios of 1:1. For each array design, two slides were hybridized with the test samples reverse-labeled in the labeling reaction. This provided for about four duplicate measurements for each clone, two of one color and two of the other, for each sample.

The differential expression assay was performed by mixing equal amounts of probes from tumor cells and normal cells of the same patient ("matched") or from tumor cells and normal cells of different patients ("unmatched") (*i.e.,* the tumor cells are from one patient and the normal cells are from a different patient). The arrays were prehybridized by incubation for about 2 hrs at 60°C in 5X SSC/0.2% SDS/1 mM EDTA, and then washed three times in water and twice in isopropanol. Following prehybridization of the array, the probe mixture was then hybridized to the array under conditions of high stringency (overnight at 42°C in 50% formamide, 5X SSC, and 0.2% SDS. After hybridization, the array was washed at 55°C three times as follows: 1) first wash in 1X SSC/0.2% SDS; 2) second wash in 0.1X SSC/0.2% SDS; and 3) third wash in 0.1X SSC.

The arrays were then scanned for green and red fluorescence using a Molecular Dynamics Generation III dual color laser-scanner/detector. The images were processed using BioDiscovery Autogene software, and the data from each scan set normalized to provide for a ratio of expression relative to normal. Data from the microarray experiments was analyzed according to the algorithms described in U.S. application serial no. 60/252,358, filed November 20, 2000, by E.J. Moler, M.A. Boyle, and F.M. Randazzo, and entitled "Precision and accuracy in cDNA microarray data," which application is specifically incorporated herein by reference.

The experiment was repeated, this time labeling the two probes with the opposite color in order to perform the assay in both "color directions." Each experiment was sometimes repeated with two more slides (one in each color direction). The level fluorescence for each sequence on the array expressed as a ratio of the geometric mean of 8 replicate spots/genes from the four arrays or 4 replicate spots/gene from 2 arrays or some other permutation. The data were normalized using the spiked positive controls present in each duplicated area, and the precision of this normalization was included in the final determination of the significance of each differential. The fluorescent intensity of each spot was also compared to the negative controls in each duplicated area to determine which spots have detected significant expression levels in each sample.

A statistical analysis of the fluorescent intensities was applied to each set of duplicate spots to assess the precision and significance of each differential measurement, resulting in a p-value testing the null hypothesis that there is no differential in the expression level between the tumor and normal samples of each patient in matched samples or between tumor and normal samples of tissue from different patients in unmatched samples. During initial analysis of the microarrays, the hypothesis was accepted if $p > 10^{-3}$, and the differential ratio was set to 1.000 for those spots. All other spots have a significant difference in expression between the tumor and normal sample. If the tumor sample has detectable expression and the normal does not, the ratio is truncated at 1000 since the value for expression in the normal sample would be zero, and the ratio would not be a mathematically useful value (*e.g.*, infinity). If the normal sample has detectable expression and the tumor does not, the ratio is truncated to 0.001, since the value for expression in the tumor sample would be zero and the ratio would not be a mathematically useful value. These latter two situations are referred to herein as "on/off." Database tables were populated using a 95% confidence level ($p>0.05$).

Table 13 (inserted prior to claims) provides the results for gene products expressed by at least 2-fold or greater in cancerous prostate, colon, or breast tissue samples relative to normal tissue samples in at least 20% of the patients tested. Table 13 includes: 1) the SEQ ID NO ("SEQ ID") assigned to each sequence for use in the present specification; 2) the sequence name ("SEQ NAME") used as an internal identifier of the sequence; 3) the name assigned to the clone from which the sequence was isolated ("CLONE ID"); 4) the percentage of patients tested in which expression levels (*e.g.*, as message level) of the gene was at least 2-fold greater in cancerous breast tissue than in matched normal tissue ("BREAST PATIENTS >=2x"); 5) the breast number ratios, indicating the number of patients upon which the provided ratio using matched breast tissue was based ("BREAST NUM RATIOS"); 6) the percentage of patients tested in which expression levels (*e.g.*, as message level) of the gene was at least 2-fold greater in cancerous colon tissue than in matched normal tissue ("COLON PATIENTS >=2x"); 7) the colon number ratios, indicating the number of patients upon which the provided ratio using matched colon tissue was based ("COLON NUM RATIOS"); 8) the percentage of patients tested in which expression levels (*e.g.*, as message level) of the gene was at least 2-fold greater in cancerous colon tissue than in unmatched normal tissue ("COLON UM >=2x"); 9) the unmatched colon number ratios, indicating the number of patients upon which the provided ratio using unmatched colon tissue was based ("COLON UM NUM RATIOS").

Table 16 (inserted prior to claims) provides the results for other gene products expressed by at least 2-fold or greater in cancerous prostate, colon, or breast tissue samples, which may be metastasized cancer samples, relative to normal tissue samples in at least 20% of the patients tested. For each set of data (i.e., the percentage of patients in which a particular sequence is up-regulated in a cancer tissue) the number of patients (Colon Cancer Patients; Colon Unmatched Met Patients and Colon Match Met Patients) is shown. If a sample is matched, it is matched to a sample from the same patient, if a sample is unmatched, the results obtained from that sample are compared to a pooled sample of an appropriate tissue type from the patients. If a sample is not from a metastasized tissue, it is from a primary tumor.

These data provide evidence that the genes represented by the polynucleotides having the indicated sequences are differentially expressed in breast, prostate, cancer as compared to normal non-cancerous breast tissue and are differentially expressed in colon cancer as compared to normal non-cancerous colon tissue

The above methods can be performed to identify genes differentially expressed in cancerous and normal cells of any type of tissue, such as prostate, lung, colon, breast, and the like.

Example 7: Antisense Regulation of Gene Expression

The expression of the differentially expressed genes represented by the polynucleotides in the cancerous cells can be further analyzed using antisense knockout technology to confirm the role and function of the gene product in tumorigenesis, *e.g.*, in promoting a metastatic phenotype.

Methods for analysis using antisense technology are well known in the art. For example, a number of different oligonucleotides complementary to the mRNA generated by the differentially expressed genes identified herein can be designed as antisense oligonucleotides, and tested for their ability to suppress expression of the genes. Sets of antisense oligomers specific to each candidate target are designed using the sequences of the polynucleotides corresponding to a differentially expressed gene and the software program HYBsimulator Version 4 (available for Windows 95/Windows NT or for Power Macintosh, RNAture, Inc. 1003 Health Sciences Road, West, Irvine, CA 92612 USA). Factors considered when designing antisense oligonucleotides include: 1) the The expression of the differentially expressed genes represented by the polynucleotides in the cancerous cells can be analyzed using antisense knockout technology to confirm the role and function of the gene product in tumorigenesis, *e.g.*, in promoting a metastatic phenotype.

A number of different oligonucleotides complementary to the mRNA generated by the differentially expressed genes identified herein can be designed as potential antisense oligonucleotides, and tested for their ability to suppress expression of the genes. Sets of antisense oligomers specific to each candidate target are designed using the sequences of the polynucleotides corresponding to a differentially expressed gene and the software program HYBsimulator Version 4 (available for Windows 95/Windows NT or for Power Macintosh, RNAture, Inc. 1003 Health Sciences Road, West, Irvine, CA 92612 USA). Factors that are considered when designing antisense oligonucleotides include: 1) the secondary structure of oligonucleotides; 2) the secondary structure of the target gene; 3) the specificity with no or minimum cross-hybridization to other expressed genes; 4) stability; 5) length and 6) terminal GC content. The antisense oligonucleotide is designed so that it will hybridize to its target sequence under conditions of high stringency at physiological temperatures (*e.g.*, an optimal temperature for the cells in culture to provide for hybridization in the cell, e.g., about 37°C), but with minimal formation of homodimers.

Using the sets of oligomers and the HYBsimulator program, three to ten antisense oligonucleotides and their reverse controls are designed and synthesized for each candidate mRNA transcript, which transcript is obtained from the gene corresponding to the target polynucleotide sequence of interest. Once synthesized and quantitated, the oligomers are screened for efficiency of a transcript knock-out in a panel of cancer cell lines. The efficiency of the knock-out is determined by analyzing mRNA levels using lightcycler quantification. The oligomers that resulted in the highest level of transcript knock-out, wherein the level was at least about 50%, preferably about 80-90%, up to 95% or more up to undetectable message, are selected for use in a cell-based proliferation assay, an anchorage independent growth assay, and an apoptosis assay.

The ability of each designed antisense oligonucleotide to inhibit gene expression is tested through transfection into LNCaP, PC3, 22Rv1, MDA-PCA-2b, or DU145 prostate carcinoma cells. For each transfection mixture, a carrier molecule (such as a lipid, lipid derivative, lipid-like molecule, cholesterol, cholesterol derivative, or cholesterol-like molecule) is prepared to a working concentration of 0.5 mM in water, sonicated to yield a uniform solution, and filtered through a 0.45 μm PVDF membrane. The antisense or control oligonucleotide is then prepared to a working concentration of 100 μM in sterile Millipore water. The oligonucleotide is further diluted in OptiMEM™ (Gibco/BRL), in a microfuge tube, to 2 μM, or approximately 20 μg oligo/ml of OptiMEM™. In a separate microfuge tube, the carrier molecule, typically in the amount of about 1.5-2 nmol carrier/μg antisense oligonucleotide, is diluted into the same volume of OptiMEM™ used to dilute the oligonucleotide. The diluted antisense oligonucleotide is immediately added to the diluted carrier and mixed by pipetting up and down. Oligonucleotide is added to the cells to a final concentration of 30 nM.

The level of target mRNA that corresponds to a target gene of interest in the transfected cells is quantitated in the cancer cell lines using the Roche LightCycler™ real-time PCR machine. Values for the target mRNA are normalized versus an internal control (*e.g.*, beta-actin). For each 20 μl reaction, extracted RNA (generally 0.2-1 μg total) is placed into a sterile 0.5 or 1.5 ml microcentrifuge tube, and water is added to a total volume of 12.5 μl. To each tube is added 7.5 μl of a buffer/enzyme mixture, prepared by mixing (in the order listed) 2.5 μl $H_2O$, 2.0 μl 10X reaction buffer, 10 μl oligo dT (20 pmol), 1.0 μl dNTP mix (10 mM each), 0.5 μl RNAsin® (20u) (Ambion, Inc., Hialeah, FL), and 0.5 μl MMLV reverse transcriptase (50u) (Ambion, Inc.). The contents are mixed by pipetting up and down, and the reaction mixture is incubated at 42°C for 1 hour. The contents of each tube are centrifuged prior to amplification.

An amplification mixture is prepared by mixing in the following order: 1X PCR buffer II, 3 mM $MgCl_2$, 140 µM each dNTP, 0.175 pmol each oligo, 1:50,000 dil of SYBR® Green, 0.25 mg/ml BSA, 1 unit *Taq* polymerase, and $H_2O$ to 20 µl. (PCR buffer II is available in 10X concentration from Perkin-Elmer, Norwalk, CT). In 1X concentration it contains 10 mM Tris pH 8.3 and 50 mM KCl. SYBR® Green (Molecular Probes, Eugene, OR) is a dye which fluoresces when bound to double stranded DNA. As double stranded PCR product is produced during amplification, the fluorescence from SYBR® Green increases. To each 20 µl aliquot of amplification mixture, 2 µl of template RT is added, and amplification is carried out according to standard protocols. The results are expressed as the percent decrease in expression of the corresponding gene product relative to non-transfected cells, vehicle-only transfected (mock-transfected) cells, or cells transfected with reverse control oligonucleotides.

Example 8: Effect of Expression on Proliferation

The effect of gene expression on the inhibition of cell proliferation can be assessed in metastatic breast cancer cell lines (MDA-MB-231 ("231")); SW620 colon colorectal carcinoma cells; SKOV3 cells (a human ovarian carcinoma cell line); or LNCaP, PC3, 22Rv1, MDA-PCA-2b, or DU145 prostate cancer cells.

Cells are plated to approximately 60-80% confluency in 96-well dishes. Antisense or reverse control oligonucleotide is diluted to 2 µM in OptiMEM™. The oligonucleotide-OptiMEM™ can then be added to a delivery vehicle, which delivery vehicle can be selected so as to be optimized for the particular cell type to be used in the assay. The oligo/delivery vehicle mixture is then further diluted into medium with serum on the cells. The final concentration of oligonucleotide for all experiments can be about 300 nM.

Antisense oligonucleotides are prepared as described above (see Example 3). Cells are transfected overnight at 37°C and the transfection mixture is replaced with fresh medium the next morning. Transfection is carried out as described above in Example 8.

Those antisense oligonucleotides that result in inhibition of proliferation of SW620 cells indicate that the corresponding gene plays a role in production or maintenance of the cancerous phenotype in cancerous colon cells. Those antisense oligonucleotides that inhibit proliferation in SKOV3 cells represent genes that play a role in production or maintenance of the cancerous phenotype in cancerous breast cells. Those antisense oligonucleotides that result in inhibition of proliferation of MDA-MB-231 cells indicate that the corresponding gene plays a role in production or maintenance of the cancerous phenotype in cancerous ovarian cells. Those antisense oligonucleotides that inhibit proliferation in LNCaP, PC3, 22Rv1, MDA-PCA-2b, or DU145 cells represent genes that play a role in production or maintenance of the cancerous phenotype in cancerous prostate cells.

Using the following antisense oligonucleotides: TTGGTTCCCAAGACAAGCCGTGAC (SEQ ID NO:1543); TCTCAACGCTACCAGGCACTCCTTG (SEQ ID NO:1544); GCACAGCCCAAAGTCAAAGGCATTA (SEQ ID NO:1545); CAGGCACTCCTTGGTCAAATGTGGG (SEQ ID NO:1546); GGACAGGGAAAGGAGAGGCTAGTCA (SEQ ID NO:1547) and TGCATTCTCTCCCACATCTCAACGC SEQ ID NO:1548, corresponding to a glutothione transferase omega identified by SEQ ID NOS: 1377 and 1541 (Chiron Candidate Id 21), were used to inhibit proliferation of SW620 colon colorectal carcinoma cells. These antisense molecules reduced glutothione transferase omega RNA expression by approximately 90%.

Example 9: Effect of Gene Expression on Cell Migration

The effect of gene expression on the inhibition of cell migration can be assessed in LNCaP, PC3, 22Rv1, MDA-PCA-2b, or DU145 prostate cancer cells using static endothelial cell binding assays, non-static endothelial cell binding assays, and transmigration assays.

For the static endothelial cell binding assay, antisense oligonucleotides are prepared as described above (see Example 8). Two days prior to use, prostate cancer cells (CaP) are plated and transfected with antisense oligonucleotide as described above (see Examples 3 and 4). On the day before use, the medium is replaced with fresh medium, and on the day of use, the medium is replaced with fresh medium containing 2 μM CellTracker green CMFDA (Molecular Probes, Inc.) and cells are incubated for 30 min. Following incubation, CaP medium is replaced with fresh medium (no CMFDA) and cells are incubated for an additional 30-60 min. CaP cells are detached using CMF PBS/2.5 mM EDTA or trypsin, spun and resuspended in DMEM/1% BSA/ 10 mM HEPES pH 7.0. Finally, CaP cells are counted and resuspended at a concentration of $1x10^6$ cells/ml.

Endothelial cells (EC) are plated onto 96-well plates at 40-50% confluence 3 days prior to use. On the day of use, EC are washed 1X with PBS and 50λ DMDM/1%BSA/10mM HEPES pH 7 is added to each well. To each well is then added 50K (50λ) CaP cells in DMEM/1% BSA/ 10mM HEPES pH 7. The plates are incubated for an additional 30 min and washed 5X with PBS containing $Ca^{++}$ and $Mg^{++}$. After the final wash, 100 μL PBS is added to each well and fluorescence is read on a fluorescent plate reader (Ab492/Em 516 nm).

For the non-static endothelial cell binding assay, CaP are prepared as described above. EC are plated onto 24-well plates at 30-40% confluence 3 days prior to use. On the day of use, a subset of EC are treated with cytokine for 6 hours then washed 2X with PBS. To each well is then added 150-200K CaP cells in DMEM/1% BSA/ 10mM HEPES pH 7. Plates are placed on a rotating shaker (70 RPM) for 30 min and then washed 3X with PBS containing $Ca^{++}$ and $Mg^{++}$. After the final wash, 500 μL PBS is added to each well and fluorescence is read on a fluorescent plate reader (Ab492/Em 516 nm).

For the transmigration assay, CaP are prepared as described above with the following changes. On the day of use, CaP medium is replaced with fresh medium containing 5 μM CellTracker green CMFDA (Molecular Probes, Inc.) and cells are incubated for 30 min. Following incubation, CaP medium is replaced with fresh medium (no CMFDA) and cells are incubated for an additional 30-60 min. CaP cells are detached using CMF PBS/2.5 mM EDTA or trypsin, spun and resuspended in EGM-2-MV medium. Finally, CaP cells are counted and resuspended at a concentration of $1x10^6$ cells/ml.

EC are plated onto FluorBlok transwells (BD Biosciences) at 30-40% confluence 5-7 days before use. Medium is replaced with fresh medium 3 days before use and on the day of use. To each transwell is then added 50K labeled CaP. 30 min prior to the first fluorescence reading, 10 μg of FITC-dextran (10K MW) is added to the EC plated filter. Fluorescence is then read at multiple time points on a fluorescent plate reader (Ab492/Em 516 nm).

Those antisense oligonucleotides that result in inhibition of binding of LNCaP, PC3, 22Rv1, MDA-PCA-2b, or DU145 prostate cancer cells to endothelial cells indicate that the corresponding gene plays a role in the production or maintenance of the cancerous phenotype in cancerous prostate cells. Those antisense oligonucleotides that result in inhibition of endothelial cell transmigration by LNCaP, PC3, 22Rv1, MDA-PCA-2b, or DU145 prostate cancer cells indicate that the corresponding gene plays a role in the production or maintenance of the cancerous phenotype in cancerous prostate cells.

Example 10: Effect of Gene Expression on Colony Formation

The effect of gene expression upon colony formation of SW620 cells, SKOV3 cells, MD-MBA-231 cells, LNCaP cells, PC3 cells, 22Rv1 cells, MDA-PCA-2b cells, and DU145 cells can be tested in a soft agar assay. Soft agar assays are conducted by first establishing a bottom layer of 2 ml of 0.6% agar in media plated fresh within a few hours of layering on the cells. The cell layer is formed on the bottom layer by removing cells transfected as described above from plates using 0.05% trypsin and washing twice in media. The cells are counted in a Coulter counter, and resuspended to $10^6$ per ml in media. 10 μl aliquots are placed with media in 96-well plates (to check counting with WST1), or diluted further for the soft agar assay. 2000 cells are plated in 800 μl 0.4% agar in duplicate wells above 0.6% agar bottom layer. After the cell layer agar solidifies, 2 ml of media is dribbled on top and antisense or reverse control oligo (produced as described in Example 8) is added without delivery vehicles. Fresh media and oligos are added every 3-4 days. Colonies form in 10 days to 3 weeks. Fields of colonies are counted by eye. Wst-1 metabolism values can be used to compensate for small differences in starting cell number. Larger fields can be scanned for visual record of differences.

Those antisense oligonucleotides that result in inhibition of colony formation of SW620 cells indicate that the corresponding gene plays a role in production or maintenance of the cancerous phenotype in cancerous colon cells. Those antisense oligonucleotides that inhibit colony formation in SKOV3 cells represent genes that play a role in production or maintenance of the cancerous phenotype in cancerous breast cells. Those antisense oligonucleotides that result in inhibition of colony formation of MDA-MB-231 cells indicate that the corresponding gene plays a role in production or maintenance of the cancerous phenotype in cancerous ovarian cells. Those antisense oligonucleotides that inhibit colony formation in LNCaP, PC3, 22Rv1, MDA-PCA-2b, or DU145 cells represent genes that play a role in production or maintenance of the cancerous phenotype in cancerous prostate cells.

Example 11: Induction of Cell Death upon Depletion of Polypeptides by Depletion of mRNA ("Antisense Knockout")

In order to assess the effect of depletion of a target message upon cell death, LNCaP, PC3, 22Rv1, MDA-PCA-2b, or DU145 cells, or other cells derived from a cancer of interest, can be transfected for proliferation assays. For cytotoxic effect in the presence of cisplatin (cis), the same protocol is followed but cells are left in the presence of 2 μM drug. Each day, cytotoxicity is monitored by measuring the amount of LDH enzyme released in the medium due to membrane damage. The activity of LDH is measured using the Cytotoxicity Detection Kit from Roche Molecular Biochemicals. The data is provided as a ratio of LDH released in the medium vs. the total LDH present in the well at the same time point and treatment (rLDH/tLDH). A positive control using antisense and reverse control oligonucleotides for BCL2 (a known anti-apoptotic gene) is included; loss of message for BCL2 leads to an increase in cell death compared with treatment with the control oligonucleotide (background cytotoxicity due to transfection).

Example 12: Functional Analysis of Gene Products Differentially Expressed in Cancer

The gene products of sequences of a gene differentially expressed in cancerous cells can be further analyzed to confirm the role and function of the gene product in tumorigenesis, *e.g.*, in promoting or inhibiting development of a metastatic phenotype. For example, the function of gene products corresponding to genes identified herein can be assessed by blocking function of the gene products in the cell. For example, where the gene product is secreted or associated with a cell surface membrane, blocking antibodies can be generated and added to cells to examine the effect upon the cell phenotype in the context of, for example, the transformation of the cell to a cancerous, particularly a metastatic, phenotype. In order to generate antibodies, a clone corresponding to a selected gene product is selected, and a sequence that represents a partial or complete coding sequence is obtained. The resulting clone is expressed, the polypeptide produced isolated, and antibodies generated. The antibodies are then combined with cells and the effect upon tumorigenesis assessed.

Where the gene product of the differentially expressed genes identified herein exhibits sequence homology to a protein of known function (*e.g.*, to a specific kinase or protease) and/or to a protein family of known function (*e.g.*, contains a domain or other consensus sequence present in a protease family or in a kinase family), then the role of the gene product in tumorigenesis, as well as the activity of the gene product, can be examined using small molecules that inhibit or enhance function of the corresponding protein or protein family.

Additional functional assays include, but are not necessarily limited to, those that analyze the effect of expression of the corresponding gene upon cell cycle and cell migration. Methods for performing such assays are well known in the art.

Example 13: Deposit Information.

Deposits of the biological materials in the tables referenced below were made with either the Agricultural Research Service Culture Collection (NRRL), 1815 North University Street, Peoria, Illinois 61604, or with the American Type Culture Collection (ATCC), 10801 University Blvd., Manasas, VA 20110-2209, under the provisions of the Budapest Treaty, on or before the filing date of the present application. The accession number indicated is assigned after successful viability testing, and the requisite fees were paid. Access to said cultures will be available during pendency of the patent application to one determined by the Commissioner to be entitled to such under 37 C.F.R. §1.14 and 35 U.S.C. §122. All restriction on availability of said cultures to the public will be irrevocably removed upon the granting of a patent based upon the application. Moreover, the designated deposits will be maintained for a period of thirty (30) years from the date of deposit, or for five (5) years after the last request for the deposit; or for the enforceable life of the U.S. patent, whichever is longer. Should a culture become nonviable or be inadvertently destroyed, or, in the case of plasmid-containing strains, lose its plasmid, it will be replaced with a viable culture(s) of the same taxonomic description.

These deposits are provided merely as a convenience to those of skill in the art, and are not an admission that a deposit is required. A license may be required to make, use, or sell the deposited materials, and no such license is hereby granted. The deposit below was received by the ATCC on or before the filing date of the present application.

**Table 14.** Cell Lines Deposited with ATCC

| Cell Line | Deposit Date | ATCC Accession No. | CMCC Accession No. |
|---|---|---|---|
| KM12L4-A | March 19, 1998 | CRL-12496 | 11606 |
| Km12C | May 15, 1998 | CRL-12533 | 11611 |
| MDA-MB-231 | May 15, 1998 | CRL-12532 | 10583 |
| MCF-7 | October 9, 1998 | CRL-12584 | 10377 |

In addition, pools of selected clones, as well as libraries containing specific clones, were assigned an "ES" number and a "CMCC" number (both internal references) and deposited with the NRRL. Table 15 (inserted before the claims) provides the NRRL Accession Nos. of the clones deposited as librarires named ES219-ES225 (CMCC5471-CMCC5477, respectively) on November 1, 2001, and of the clones deposited as a library named ES226 (CMCC5478) on November 7, 2001.

Retrieval of Individual Clones from Deposit of Pooled Clones. Where the biological deposit is composed of a pool of cDNA clones or a library of cDNA clones, the deposit was prepared by first transfecting each of the clones into separate bacterial cells. The clones in the pool or library were then deposited as a pool of equal mixtures in the composite deposit. Particular clones can be obtained from the composite deposit using methods well known in the art. For example, a bacterial cell containing a particular clone can be identified by isolating single colonies, and identifying colonies containing the specific clone through standard colony hybridization techniques, using an oligonucleotide probe or probes designed to specifically hybridize to a sequence of the clone insert (*e.g.*, a probe based upon unmasked sequence of the encoded polynucleotide having the indicated SEQ ID NO). The probe should be designed to have a $T_m$ of approximately 80°C (assuming 2°C for each A or T and 4°C for each G or C). Positive colonies can then be picked, grown in culture, and the recombinant clone isolated. Alternatively, probes designed in this manner can be used to PCR to isolate a nucleic acid molecule from the pooled clones according to methods well known in the art, *e.g.*, by purifying the cDNA from the deposited culture pool, and using the probes in PCR reactions to produce an amplified product having the corresponding desired polynucleotide sequence.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it is readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims. Those skilled in the art will recognize, or be able to ascertain, using not more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such specific embodiments and equivalents are intended to be encompassed by the following claims.

**Table 8**

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 1 | 3538.O24.GZ43_504925 | AF047717 | Streptomyces chrysomallus actinomycin synthetase II (acmB) gene, complete cds | 1.17E-04 |
| 2 | 3538.P11.GZ43_504718 | AF111848 | Homo sapiens PRO0529 mRNA, complete cds | 2.00E-06 |
| 3 | 3541.A04.GZ43_504975 | X58178 | S.pyogenes for emm41 gene | 5.00E-06 |
| 4 | 3541.A05.GZ43_504991 | AF190638 | Mus musculus nephrin NPHS1 (Nphs1) gene, partial cds | 2.00E-06 |
| 5 | 3541.A16.GZ43_505167 | AB024689 | Mus musculus gene, exon 3, partial sequence | 6.00E-06 |
| 6 | 3541.A23.GZ43_505279 | M14155 | Human insulin-like growth factor (IGF-I) IB gene, exon 4 | 3.00E-06 |
| 7 | 3541.B04.GZ43_504976 | U32801 | Haemophilus influenzae Rd section 116 of 163 of the complete genome | 1.10E-05 |
| 8 | 3541.B17.GZ43_505184 | X89398 | H.sapiens ung gene for uracil DNA-glycosylase | 1.21E-04 |
| 9 | 3538.G08.GZ43_504661 | AF270390 | Staphylococcus epidermidis strain SR1 clone step.4045d08 genomic sequence | 3.00E-06 |
| 10 | 3538.G17.GZ43_504805 | AC006623 | Caenorhabditis elegans clone C52E2, complete sequence | 4.00E-06 |
| 11 | 3538.G19.GZ43_504837 | AB042425 | Homo sapiens Pim-2h, hUGT2, hUGT1, genes for pim-2 protooncogene homolog, UDP-galactose transporter 1, UDP-galactose transporter 2, complete cds | 6.60E-11 |
| 12 | 3538.G22.GZ43_504885 | L08338 | Human immunodeficiency virus type 1 proviral envelope glycoprotein gene V3 region from A196/4537, clone 6 (from adult) | 3.10E-07 |
| 13 | 3538.H05.GZ43_504614 | AE006731 | Sulfolobus solfataricus section 90 of 272 of the complete genome | 2.00E-06 |
| 14 | 3538.H21.GZ43_504870 | AL121807 | S.pombe chromosome III cosmid c132 | 1.30E-05 |
| 15 | 3538.I08.GZ43_504663 | AF186379 | Homo sapiens ligand effect modulator-6 (LEM6) mRNA, complete cds | 8.00E-10 |
| 16 | 3538.I13.GZ43_504743 | AC007658 | Arabidopsis thaliana chromosome II section 216 of 255 of the complete sequence. Sequence from clones F27I1 | 3.30E-08 |
| 17 | 3538.J22.GZ43_504888 | X04616 | Anacystis nidulans R2 psbAI gene for photosystem II Q(B) protein | 8.90E-07 |
| 18 | 3538.K12.GZ43_504729 | X91656 | M.musculus Srp20 gene | 4.40E-05 |
| 19 | 3538.K23.GZ43_504905 | M62849 | Human papillomavirus ORFs | 4.40E-07 |
| 20 | 3538.L16.GZ43_504794 | AE001382 | Plasmodium falciparum chromosome 2, section 19 of 73 of the complete sequence | 7.00E-06 |
| 21 | 3538.M02.GZ43_504571 | U07976 | Human T cell receptor beta (TCRBV7S2, TCRBV13S2-1, TCRBV6S7-1) genes, TCRBV deleted 2 haplotype, partial cds | 7.00E-06 |
| 22 | 3538.M05.GZ43_504619 | AC079878 | Homo sapiens BAC clone RP11-343P21 from 7, complete sequence | 1.40E-07 |
| 23 | 3538.M08.GZ43_504667 | AF182668 | Zenaida galapagoensis beta-fibrinogen gene, partial sequence | 4.70E-08 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 24 | 3538.N20.GZ43_504860 | AB033411 | Taenia crassiceps mitochondrial gene for cytochrome c oxidase subunit 1, partial cds | 6.80E-07 |
| 25 | 3538.O07.GZ43_504653 | X68019 | Feline Immunodeficiency Virus GAG gene | 4.00E-06 |
| 26 | 3541.E11.GZ43_505091 | M73447 | Human repeat polymorphism at locus D9S59 | 3.00E-08 |
| 27 | 3541.E14.GZ43_505139 | AJ243419 | Acaulospora trappei partial 18S rRNA, 5.8S rRNA and partial 28S rRNA genes and internal transcribed spacers 1 and 2 (ITS1, ITS2), isolate AU 219 | 1.10E-07 |
| 28 | 3541.E15.GZ43_505155 | U13679 | Human lactate dehydrogenase-A (LDH-A) gene, promoter region | 3.50E-10 |
| 29 | 3541.G17.GZ43_505189 | AE004851 | Pseudomonas aeruginosa PA01, section 412 of 529 of the complete genome | 1.30E-05 |
| 30 | 3541.H14.GZ43_505142 | AJ252202 | Drosophila melanogaster D-COQ7 gene for putative COQ7 isologue, exons 1-3 | 9.00E-06 |
| 31 | 3541.I15.GZ43_505159 | X98371 | D.subobscura sex-lethal gene | 6.00E-06 |
| 32 | 3541.I17.GZ43_505191 | AK023918 | Homo sapiens cDNA FLJ13856 fis, clone THYRO1000988 | 1.70E-22 |
| 33 | 3541.I18.GZ43_505207 | AF329081 | Bos taurus AMP-activated protein kinase gamma-1 (PRKAG1) gene, partial cds | 5.30E-33 |
| 34 | 3541.J19.GZ43_505224 | AF002749 | Psychotria urceolata ribosomal protein S16 (rps16) gene, chloroplast gene encoding chloroplast protein, partial intron | 3.01E-03 |
| 35 | 3541.K09.GZ43_505065 | AF027607 | Gallus gallus L-type voltage-gated calcium channel alpha1D subunit ChCaChA1D precursor mRNA, complete intron sequence | 9.00E-06 |
| 36 | 3541.L19.GZ43_505226 | AE003949 | Xylella fastidiosa 9a5c, section 95 of 229 of the complete genome | 2.00E-06 |
| 37 | 3541.M02.GZ43_504955 | BC004556 | Homo sapiens, Similar to CG7083 gene product, clone MGC:10534 IMAGE:3957147, mRNA, complete cds | 6.20E-07 |
| 38 | 3541.M07.GZ43_505035 | X05616 | Kangaroo rat repetitive DNA with insertion sequence | 4.80E-08 |
| 39 | 3541.M18.GZ43_505211 | M81888 | Parvovirus LuII DNA sequence | 6.60E-05 |
| 40 | 3541.O04.GZ43_504989 | AF081828 | Ixodes hexagonus mitochondrial DNA, complete genome | 3.00E-06 |
| 41 | 3541.O13.GZ43_505133 | AK026465 | Homo sapiens cDNA: FLJ22812 fis, clone KAIA2955 | 8.00E-06 |
| 42 | 3541.O23.GZ43_505293 | X54859 | Porcine TNF-alpha and TNF-beta genes for tumour necrosis factors alpha and beta, respectively | 2.90E-05 |
| 43 | 3541.P05.GZ43_505006 | AE006642 | Sulfolobus solfataricus section 1 of 272 of the complete genome | 3.50E-05 |
| 44 | 3541.P22.GZ43_505278 | U10400 | Saccharomyces cerevisiae chromosome VIII cosmid L2825 | 1.80E-05 |
| 45 | 3544.A09.GZ43_505439 | X75677 | C.parapsilosis mt tRNA genes (591bps) | 3.70E-08 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 46 | 3544.A13.GZ43_505503 | D28811 | Schistosoma japonicum mRNA for paramyosin, complete cds | 5.40E-05 |
| 47 | 3544.A14.GZ43_505519 | M87111 | Human immunodeficiency virus type 2 (FORTC2) reverse transcriptase fragment | 2.90E-05 |
| 48 | 3544.A17.GZ43_505567 | L23650 | Caenorhabditis elegans cosmid C27D11, complete sequence | 5.60E-07 |
| 49 | 3544.B02.GZ43_505328 | AF060543 | Homo sapiens importin alpha 7 subunit mRNA, complete cds | 1.50E-49 |
| 50 | 3544.B09.GZ43_505440 | AB051473 | Homo sapiens mRNA for KIAA1686 protein, partial cds | 1.80E-05 |
| 51 | 3544.B18.GZ43_505584 | AJ224821 | Loxodonta africana complete mitochondrial genomic sequence | 4.00E-06 |
| 52 | 3544.E05.GZ43_505379 | AL451187 | Human DNA sequence from clone RP11-49J23 on chromosome 6, complete sequence [Homo sapiens] | 1.30E-07 |
| 53 | 3544.E18.GZ43_505587 | L08338 | Human immunodeficiency virus type 1 proviral envelope glycoprotein gene V3 region from A196/4537, clone 6 (from adult) | 3.30E-07 |
| 54 | 3544.F06.GZ43_505396 | X60833 | R.norvegicus TDO2 gene for tryptophan 2,3-dioxygenase, exon 6 | 7.80E-07 |
| 55 | 3544.F16.GZ43_505556 | U72716 | Drosophila melanogaster D3-100EF mRNA, complete cds | 2.00E-06 |
| 56 | 3544.G06.GZ43_505397 | AC002359 | Homo sapiens Xp22 Cosmid U239B3 (from Lawrence Livermore X library) complete sequence | 1.60E-05 |
| 57 | 3544.G10.GZ43_505461 | X56015 | Crithidia oncopelti mitochondrial ND4, ND5, COI, 12S ribosomal RNA genes for NADH dehydrogenase subunit 4/5, cytochrome oxidase subunit I and 12S ribosomal RNA | 4.80E-05 |
| 58 | 3544.G11.GZ43_505477 | U80927 | Dictyostelium discoideum unknown protein gene, complete cds | 9.00E-08 |
| 59 | 3544.G12.GZ43_505493 | AF245483 | Oryza sativa OSE4 (OSE4) gene, complete cds | 1.70E-07 |
| 60 | 3544.H03.GZ43_505350 | Y12855 | Homo sapiens P2X7 gene, exon 12 and 13 | 2.30E-05 |
| 61 | 3544.H15.GZ43_505542 | AF194829 | Tetragonia tetragonioides NADH dehydrogenase (ndhF) gene, partial cds; chloroplast gene for chloroplast product | 2.00E-06 |
| 62 | 3544.H24.GZ43_505686 | BC008353 | Homo sapiens, Similar to RIKEN cDNA 0610008P16 gene, clone MGC:15937 IMAGE:3537224, mRNA, complete cds | 2.50E-18 |
| 63 | 3544.I07.GZ43_505415 | AF010533 | Plasmodium falciparum microsatellite TA21 sequence | 1.80E-08 |
| 64 | 3544.I15.GZ43_505543 | D29794 | Mouse gene for T cell receptor gamma chain | 3.00E-06 |
| 65 | 3544.I20.GZ43_505623 | AE000677 | Aquifex aeolicus section 9 of 109 of the complete genome | 4.00E-06 |
| 66 | 3544.J04.GZ43_505368 | Z97015 | Lactococcus lactis cremoris sucrose gene cluster | 1.00E-06 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 67 | 3544.J11.GZ43_505480 | M67480 | Human prothymosin-alpha gene, complete cds | 5.10E-10 |
| 68 | 3544.J13.GZ43_505512 | AJ249884 | Lepeophtheirus salmonis microsatellite DNA, locus Ls.NUIG.09 | 5.70E-08 |
| 69 | 3544.J23.GZ43_505672 | AJ245823 | Trypanosoma brucei PK4 gene for protein kinase | 6.00E-06 |
| 70 | 3544.K16.GZ43_505561 | U18191 | Human HLA class I genomic survey sequence | 2.50E-07 |
| 71 | 3544.L11.GZ43_505482 | X07127 | Kluyveromyces lactis killer plasmid k1 DNA | 2.90E-05 |
| 72 | 3544.L13.GZ43_505514 | BC005028 | Homo sapiens, hypothetical protein FLJ11323, clone MGC:12582 IMAGE:3953383, mRNA, complete cds | 1.80E-31 |
| 73 | 3544.M06.GZ43_505403 | AC006687 | Caenorhabditis elegans cosmid T20C7, complete sequence | 2.30E-05 |
| 74 | 3544.M10.GZ43_505467 | M92378 | Mus musculus GABA transporter mRNA sequence | 1.30E-05 |
| 75 | 3544.N07.GZ43_505420 | U48705 | Human receptor tyrosine kinase DDR gene, complete cds | 7.40E-07 |
| 76 | 3544.N12.GZ43_505500 | BC007621 | Homo sapiens, Similar to Orthodenticle (Drosophila) homolog 1, clone MGC:15736 IMAGE:3355563, mRNA, complete cds | 5.70E-07 |
| 77 | 3544.N19.GZ43_505612 | AF270077 | Staphylococcus epidermidis strain SR1 clone step.1047c06 genomic sequence | 2.00E-07 |
| 78 | 3544.O03.GZ43_505357 | U15681 | Myrmecia pilosula HI87-156 mitochondrion cytochrome b gene, partial cds | 1.00E-06 |
| 79 | 3544.O10.GZ43_505469 | AF056032 | Homo sapiens kynurenine 3-hydroxylase mRNA, complete cds | 5.00E-06 |
| 80 | 3544.O15.GZ43_505549 | U37373 | Xenopus laevis tail-specific thyroid hormone up-regulated (gene 5) mRNA, complete cds | 3.00E-06 |
| 81 | 3544.O20.GZ43_505629 | D66906 | Bombyx mori DNA for sorbitol dehydrogenase, complete cds | 2.00E-06 |
| 82 | 3544.P18.GZ43_505598 | J04357 | Red clover necrotic mosaic virus RNA-1, complete sequence | 4.00E-06 |
| 83 | 3547.A04.GZ43_505743 | AF118558 | Mus musculus hitchhiker-3, hitchhiker-4, and hitchhiker-5 mRNA sequences | 5.40E-07 |
| 84 | 3547.A11.GZ43_505855 | U93874 | Bacillus subtilis cysteine synthase (yrhA), cystathionine gamma-lyase (yrhB), YrhC (yrhC), YrhD (yrhD), formate dehydrogenase chain A (yrhE), YrhF (yrhF), formate dehydrogenase (yrhG), YrhH (yrhH), regulatory protein (yrhI), cytochrome P450 102 (yrhJ),> | 4.00E-06 |
| 85 | 3547.A24.GZ43_506063 | AL157466 | Homo sapiens mRNA; cDNA DKFZp761E2423 (from clone DKFZp761E2423) | 8.80E-07 |
| 86 | 3547.C05.GZ43_505761 | X52589 | Bovine rotavirus RNA for virus protein 2 (VP2) | 1.00E-05 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 87 | 3547.C17.GZ43_505953 | U67594 | Methanococcus jannaschii section 136 of 150 of the complete genome | 3.80E-05 |
| 88 | 3547.C23.GZ43_506049 | AJ250862 | Bacillus sp. HIL-Y85/54728 mersacidin biosynthesis gene cluster (mrsK2, mrsR2, mrsF, mrsG, mrsE, mrsA, mrsR1, mrsD, mrsM and mrsT genes) | 1.20E-05 |
| 89 | 3547.D19.GZ43_505986 | AF050491 | Microgadus tomcod aromatic hydrocarbon receptor (ahr) gene, exons 8-11, partial cds | 4.00E-06 |
| 90 | 3547.D23.GZ43_506050 | M33190 | Rat cytochrome P450 II A3 (CYP2A3) gene, complete cds | 5.80E-05 |
| 91 | 3547.E04.GZ43_505747 | L35658 | Homo sapiens (subclone H8 9_d12 from P1 35 H5 C8) DNA sequence | 7.70E-07 |
| 92 | 3547.F02.GZ43_505716 | AF038190 | Homo sapiens clone 23582 mRNA sequence | 1.10E-07 |
| 93 | 3547.F10.GZ43_505844 | AY008833 | Staphylococcus aureus tcaR-tcaA-tcaB operon, complete sequences | 5.00E-06 |
| 94 | 3547.F20.GZ43_506004 | AB037821 | Homo sapiens mRNA for KIAA1400 protein, partial cds | 1.00E-06 |
| 95 | 3547.G02.GZ43_505717 | M88397 | Naegleria fowleri virulence-related protein (NF314) mRNA, complete cds | 3.70E-07 |
| 96 | 3547.G09.GZ43_505829 | AJ315644 | Homo sapiens mRNA for proton myo-inositol symporter (Hmit gene) | 7.90E-07 |
| 97 | 3547.G22.GZ43_506037 | Z33603 | P.radiata (Pr1.6) microsatellite DNA, 703bp | 1.70E-07 |
| 98 | 3547.H12.GZ43_505878 | L04309 | Shigella flexneri ipgD, ipgE, ipgF genes, complete cds | 3.00E-06 |
| 99 | 3547.H14.GZ43_505910 | AL137502 | Homo sapiens mRNA; cDNA DKFZp761H171 (from clone DKFZp761H171); partial cds | 2.90E-07 |
| 100 | 3547.I07.GZ43_505799 | M15332 | B.sphaericus ermG gene encoding rRNA methyltransferase (macrolide-lincosamide-streptogramin B resistance element) | 7.00E-06 |
| 101 | 3547.I16.GZ43_505943 | AF015157 | Homo sapiens clone HS19.12 Alu-Ya5 sequence | 4.70E-10 |
| 102 | 3547.I17.GZ43_505959 | AE007758 | Clostridium acetobutylicum ATCC824 section 246 of 356 of the complete genome | 3.00E-06 |
| 103 | 3547.I20.GZ43_506007 | L37606 | Medicago sativa (clone GG16-1) NADH-dependent glutamate synthase gene, complete cds | 1.50E-05 |
| 104 | 3547.J05.GZ43_505768 | Z16911 | H. sapiens (D20S113) DNA segment containing (CA) repeat; clone AFM205th8; single read | 2.80E-07 |
| 105 | 3547.J10.GZ43_505848 | Z37803 | HIV-1 DNA V3 region (patient 15, sample CSF, clone 9 ) | 8.80E-07 |
| 106 | 3547.J20.GZ43_506008 | AF013273 | Candida albicans histidine kinase 1 gene, complete cds | 3.30E-05 |
| 107 | 3547.J22.GZ43_506040 | AF289080 | Lycopersicon esculentum alpha-galactosidase gene, partial cds | 4.00E-06 |

## Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 108 | 3547.K01.GZ43_505705 | AF267863 | Homo sapiens DC43 mRNA, complete cds | 7.30E-22 |
| 109 | 3547.L09.GZ43_505834 | Z22175 | Caenorhabditis elegans cosmid K01F9, complete sequence | 1.40E-05 |
| 110 | 3547.L11.GZ43_505866 | AJ288648 | Limnodynastes tasmaniensis mitochondrial partial nadh4 gene for NADH dehydrogenase subunit 4 and partial tRNA-His gene, sample 26 from Australia:Boolara | 5.90E-07 |
| 111 | 3547.L16.GZ43_505946 | AE001293 | Chlamydia trachomatis section 20 of 87 of the complete genome | 7.10E-07 |
| 112 | 3547.L22.GZ43_506042 | AF287006 | Danio rerio T-box brain 1 mRNA, partial cds | 7.00E-06 |
| 113 | 3547.M02.GZ43_505723 | AE007788 | Clostridium acetobutylicum ATCC824 section 276 of 356 of the complete genome | 1.00E-05 |
| 114 | 3547.M07.GZ43_505803 | Z46252 | M.musculus DNA for region surrounding retrovirus restriction locus Fv1 | 6.00E-06 |
| 115 | 3547.M08.GZ43_505819 | AB020684 | Homo sapiens mRNA for KIAA0877 protein, partial cds | 1.50E-05 |
| 116 | 3547.M16.GZ43_505947 | AF335240 | Petunia x hybrida MADS-box transcription factor FBP22 (FBP22) mRNA, complete cds | 3.00E-06 |
| 117 | 3547.N06.GZ43_505788 | AF299346 | Renispora flavissima isolate CEH313 18S ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and 28S ribosomal RNA gene, partial sequence | 1.70E-08 |
| 118 | 3547.O03.GZ43_505741 | AE002344 | Chlamydia muridarum, section 72 of 85 of the complete genome | 6.60E-07 |
| 119 | 3547.O07.GZ43_505805 | D50608 | Rat gene for cholecystokinin type-A receptor (CCKAR), complete cds | 1.60E-05 |
| 120 | 3547.O14.GZ43_505917 | AL137502 | Homo sapiens mRNA; cDNA DKFZp761H171 (from clone DKFZp761H171); partial cds | 2.90E-07 |
| 121 | 3547.P18.GZ43_505982 | AJ131734 | Plasmodium berghei DNA including upstream sequence NTS and 5'ETS of the 18S rRNA gene (A rRNA gene unit) | 6.10E-07 |
| 122 | 3547.P21.GZ43_506030 | AC006619 | Caenorhabditis elegans cosmid C46C11, complete sequence | 1.70E-05 |
| 123 | 3547.P22.GZ43_506046 | AJ000871 | Streptococcus mitis comC, comD, comE genes, isolate B5 | 2.00E-06 |
| 124 | 3550.A12.GZ43_506255 | M22310 | Human epidermal growth factor receptor proto-oncogene downstream enhancer | 4.80E-07 |
| 125 | 3550.A16.GZ43_506319 | L39435 | Senecio mikanioides chloroplast NADH dehydrogenase (ndhF) gene, complete cds | 2.00E-06 |
| 126 | 3550.B06.GZ43_506160 | D14161 | Hordeum vulgare ids-4 mRNA, complete cds | 1.10E-08 |

**Table 8**

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 127 | 3550.C01.GZ43_506081 | AK025182 | Homo sapiens cDNA: FLJ21529 fis, clone COL05981 | 4.20E-09 |
| 128 | 3550.C22.GZ43_506417 | X52028 | Rattus norvegicus P450 IID3 gene | 1.41E-04 |
| 129 | 3550.D16.GZ43_506322 | Y10345 | H.sapiens GalNAc-T3 gene, 3'UTR | 5.00E-07 |
| 130 | 3550.D23.GZ43_506434 | AF134403 | Escherichia coli plasmid pAA2 Shf (shf), hexosyltransferase homolog (capU), and VirK (virK) genes, complete cds | 6.90E-07 |
| 131 | 3550.E02.GZ43_506099 | U66074 | Tritrichomonas foetus putative superoxide dismutase 2 (SOD2) gene, complete cds | 8.90E-07 |
| 132 | 3550.E06.GZ43_506163 | Z23341 | H. sapiens (D8S528) DNA segment containing (CA) repeat; clone AFM080xh7; single read | 2.30E-08 |
| 133 | 3550.F06.GZ43_506164 | M59447 | Drosophila melanogaster Sex-lethal (Sx1) mRNA, complete cds | 3.00E-06 |
| 134 | 3550.F08.GZ43_506196 | M24901 | Rabbit pulmonary surfactant-associated protein (SP-B) mRNA, complete cds | 3.40E-07 |
| 135 | 3550.F20.GZ43_506388 | AF216169 | Simicratea welwitschii clone 2 phytochrome B (PHYB) gene, exon 1 and partial cds | 5.40E-08 |
| 136 | 3550.F22.GZ43_506420 | AP000739 | Arabidopsis thaliana genomic DNA, chromosome 3, P1 clone:MEK6 | 2.20E-05 |
| 137 | 3550.G02.GZ43_506101 | AL022342 | Human DNA sequence from clone RP1-29M10 on chromosome 20, complete sequence [Homo sapiens] | 7.40E-05 |
| 138 | 3550.G08.GZ43_506197 | AK021312 | Mus musculus 13 days embryo stomach cDNA, RIKEN full-length enriched library, clone:D530039A21, full insert sequence | 3.60E-08 |
| 139 | 3550.G10.GZ43_506229 | M31684 | D.melanogaster cytoskeleton-like bicaudalD protein (BicD) mRNA, complete cds | 3.00E-06 |
| 140 | 3550.G15.GZ43_506309 | AF087141 | Mus musculus uncharacterized long terminal repeat, complete sequence; and valyl-tRNA synthetase (G7a) gene, complete cds | 4.00E-06 |
| 141 | 3550.G23.GZ43_506437 | X02547 | Trypanosoma brucei mitochondrial genes for 12S and 9S ribosomal RNA | 2.00E-06 |
| 142 | 3550.H10.GZ43_506230 | U55711 | Human ataxia-telangiectasia (ATM) gene, exon 11 | 6.10E-08 |
| 143 | 3550.H21.GZ43_506406 | Z68755 | Human DNA sequence from cosmid L118D5, Huntington's Disease Region, chromosome 4p16.3 | 2.00E-06 |
| 144 | 3550.H23.GZ43_506438 | AF151388 | Dermatobia hominis strain Alfenas tRNA-Ile gene, partial sequence; D-loop, complete sequence; and 12S ribosomal RNA, partial sequence; mitochondrial genes for mitochondrial products | 1.20E-07 |

**Table 8**

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 145 | 3550.I03.GZ43_506119 | AF117258 | Staphylococcus aureus plasmid pIP680 replication protein RepE (repE) gene, partial cds; resolvase (res), acetyltransferase Vat (vat), and hydrolase VgB (vgb) genes, complete cds; and unknown gene | 6.50E-08 |
| 146 | 3550.I19.GZ43_506375 | AE002781 | Drosophila melanogaster genomic scaffold 142000013385442, complete sequence | 3.90E-05 |
| 147 | 3550.I21.GZ43_506407 | AE001002 | Archaeoglobus fulgidus section 105 of 172 of the complete genome | 4.20E-05 |
| 148 | 3550.J05.GZ43_506152 | AF080689 | Homo sapiens protein kinase PITSLRE (CDC2L2) gene, exons 8 and 9 | 5.50E-10 |
| 149 | 3550.J11.GZ43_506248 | Z82761 | R.prowazekii genomic DNA fragment (clone A793R) | 1.00E-06 |
| 150 | 3550.K05.GZ43_506153 | X15407 | Maize pseudo-Gpa2 pseudogene for glyceraldehyde-3-phosphate dehydrogenase subunit A | 3.20E-05 |
| 151 | 3550.K09.GZ43_506217 | X62631 | S.pombe wis1 gene for protein kinase | 1.50E-07 |
| 152 | 3550.K14.GZ43_506297 | M59743 | Rabbit cardiac muscle Ca-2+ release channel (ryanodine receptor) mRNA, complete cds | 1.00E-06 |
| 153 | 3550.L16.GZ43_506330 | AF201383 | Buchnera aphidicola isopropylmalate dehydratase subunit (leuC) gene, partial cds | 1.00E-06 |
| 154 | 3550.L19.GZ43_506378 | M77244 | H.sapiens erythropoietin receptor (EPOR) gene, 5' end | 4.00E-09 |
| 155 | 3550.L23.GZ43_506442 | L76259 | Homo sapiens PTS gene, complete cds | 8.00E-06 |
| 156 | 3550.M21.GZ43_506411 | M87339 | Human replication factor C, 37-kDa subunit mRNA, complete cds | 5.00E-06 |
| 157 | 3550.N01.GZ43_506092 | AF191009 | Helicobacter pylori strain ChinaF30A cag pathogenicity island polymorphic right end, type IIIa motif | 1.10E-07 |
| 158 | 3550.N07.GZ43_506188 | AF235499 | Mus musculus SH2-containing inositol 5-phosphatase (Ship) gene, exons 3 through 6 | 1.55E-04 |
| 159 | 3550.O03.GZ43_506125 | D14813 | Human DNA for osteopontin, complete cds | 4.50E-05 |
| 160 | 3550.O04.GZ43_506141 | U08596 | Canis familiaris delayed rectifier K+ channel mRNA, partial cds | 6.00E-06 |
| 161 | 3550.O08.GZ43_506205 | XM_017044 | Homo sapiens similar to diaphanous (Drosophila, homolog) 2 (H. sapiens) (LOC91459), mRNA | 6.40E-09 |
| 162 | 3550.O15.GZ43_506317 | U15977 | Mus musculus long chain fatty acyl CoA synthetase mRNA, complete cds | 2.80E-05 |
| 163 | 3550.O17.GZ43_506349 | X62578 | C.caldarium plastid genes ompR', psbD, psbC, rps16 and groEL | 2.80E-05 |
| 164 | 3550.O18.GZ43_506365 | L34363 | Human X-linked nuclear protein (XNP) gene, complete cds | 4.00E-06 |
| 165 | 3550.O21.GZ43_506413 | AB056784 | Macaca fascicularis brain cDNA clone:QnpA-11501, full insert sequence | 5.20E-07 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 166 | 3550.P18.GZ43_506366 | AK002041 | Homo sapiens cDNA FLJ11179 fis, clone PLACE1007450 | 5.30E-07 |
| 167 | 3550.P23.GZ43_506446 | AF200361 | Rattus norvegicus cytochrome P450 4F1 (Cyp4F1) gene, complete cds | 1.40E-05 |
| 168 | 3553.A09.GZ43_506591 | AL109980 | Human DNA sequence from clone RP4-697G8 on chromosome 22, complete sequence [Homo sapiens] | 3.50E-12 |
| 169 | 3553.B07.GZ43_506560 | L37056 | Strongylocentrotus purpuratus myc protein mRNA, complete cds | 4.60E-07 |
| 170 | 3553.B16.GZ43_506704 | U43542 | Nicotiana tabacum diphenol oxidase mRNA, complete cds | 2.00E-06 |
| 171 | 3553.B22.GZ43_506800 | L34040 | Homo sapiens stromelysin gene, promoter region | 6.00E-06 |
| 172 | 3553.D04.GZ43_506514 | Y07599 | S.pombe mRNA for dmf1 gene | 9.40E-07 |
| 173 | 3553.D07.GZ43_506562 | X13835 | R.norvegicus CaMII gene, exons 3,4 & 5 | 2.00E-06 |
| 174 | 3553.D14.GZ43_506674 | L38424 | Bacillus subtilis dihydropicolinate reductase (jojE) gene, complete cds; poly(A) polymerase (jojI) gene, complete cds; biotin acetyl-CoA-carboxylase ligase (birA) gene, complete cds; jojC, jojD, jojF, jojG, jojH genes, complete cds's | 1.80E-05 |
| 175 | 3553.D19.GZ43_506754 | X53431 | Yeast gene for STE11 | 9.00E-06 |
| 176 | 3553.E08.GZ43_506579 | AF062863 | Arabidopsis thaliana putative transcription factor (MYB11) mRNA, partial cds | 1.80E-07 |
| 177 | 3553.E09.GZ43_506595 | X71067 | X.laevis XFG 5-1 and XFG 5-2 genes for zinc finger proteins | 6.60E-05 |
| 178 | 3553.F12.GZ43_506644 | X63223 | B.taurus CI-MNLL mRNA for ubiquinone oxidoreductase complex | 6.90E-08 |
| 179 | 3553.F13.GZ43_506660 | L81869 | Homo sapiens (subclone 1_c4 from P1 H55) DNA sequence, complete sequence | 3.00E-08 |
| 180 | 3553.F19.GZ43_506756 | U97190 | Caenorhabditis elegans cosmid B0025, complete sequence | 3.00E-06 |
| 181 | 3553.G05.GZ43_506533 | S76404 | beta -HKA=H,K-ATPase beta-subunit [rats, Genomic, 8983 nt, segment 2 of 2] | 8.00E-06 |
| 182 | 3553.G06.GZ43_506549 | X68048 | Phaseolus vulgaris chloroplast DNA for tRNA-His gene region | 5.00E-06 |
| 183 | 3553.G07.GZ43_506565 | AF068289 | Homo sapiens HDCMD34P mRNA, complete cds | 4.40E-12 |
| 184 | 3553.G21.GZ43_506789 | Z33603 | P.radiata (Pr1.6) microsatellite DNA, 703bp | 1.70E-07 |
| 185 | 3553.H06.GZ43_506550 | AF090901 | Homo sapiens clone HQ0195$ PRO0195 mRNA, complete cds | 8.00E-07 |
| 186 | 3553.H09.GZ43_506598 | AF270105 | Staphylococcus epidermidis strain SR1 clone step.1049c09 genomic sequence | 9.80E-07 |
| 187 | 3553.H21.GZ43_506790 | Z18359 | Glycine max seed-specific low molecular weight sulfur-rich protein | 2.00E-06 |
| 188 | 3553.I13.GZ43_506663 | AF155115 | Homo sapiens NY-REN-58 antigen mRNA, complete cds | 1.70E-07 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 189 | 3553.I16.GZ43_506711 | AF270229 | Staphylococcus epidermidis strain SR1 clone step.1055d10 genomic sequence | 1.20E-05 |
| 190 | 3553.J12.GZ43_506648 | U53400 | Rattus norvegicus chromosome 10 microsatellite sequence D10Mco21 | 1.55E-01 |
| 191 | 3553.J14.GZ43_506680 | M10014 | Homo sapiens map 4q28 fibrinogen (FGG) gene, alternative splice products, complete cds | 9.00E-06 |
| 192 | 3553.J16.GZ43_506712 | Z23341 | H. sapiens (D8S528) DNA segment containing (CA) repeat; clone AFM080xh7; single read | 2.30E-08 |
| 193 | 3553.J17.GZ43_506728 | AK021312 | Mus musculus 13 days embryo stomach cDNA, RIKEN full-length enriched library, clone:D530039A21, full insert sequence | 3.60E-08 |
| 194 | 3553.J22.GZ43_506808 | AF120279 | Mus musculus proline dehydrogenase mRNA, complete cds | 5.00E-06 |
| 195 | 3553.J24.GZ43_506840 | Z18359 | Glycine max seed-specific low molecular weight sulfur-rich protein | 2.00E-06 |
| 196 | 3553.K01.GZ43_506473 | U31465 | Kluyveromyces lactis telomerase RNA component (TER1) gene, complete sequence | 2.00E-06 |
| 197 | 3553.K02.GZ43_506489 | X60672 | M.musculus mRNA for radixin | 1.00E-06 |
| 198 | 3553.K03.GZ43_506505 | Z71943 | G.hyalina (92-89) DNA for internal transcribed spacer 1 | 1.06E-02 |
| 199 | 3553.K05.GZ43_506537 | M80596 | Saccharomyces cerevisiae VAC1 gene (required for vacuole inheritance and vacuole protein sorting), complete cds | 6.00E-06 |
| 200 | 3553.K07.GZ43_506569 | AJ275317 | Cicer arietinum partial mRNA for malate dehydrogenase | 7.60E-07 |
| 201 | 3553.K15.GZ43_506697 | X57377 | Mouse dilute myosin heavy chain gene for novel heavy chain with unique C-terminal region | 2.40E-05 |
| 202 | 3538.A11.GZ43_504703 | Z75199 | S.cerevisiae chromosome XV reading frame ORF YOR291w | 6.00E-06 |
| 203 | 3538.A24.GZ43_504911 | AF270077 | Staphylococcus epidermidis strain SR1 clone step.1047c06 genomic sequence | 2.00E-07 |
| 204 | 3538.B01.GZ43_504544 | AF368255 | Arabidopsis thaliana small zinc finger-like protein TIM13 mRNA, complete cds; nuclear gene for mitochondrial product | 4.10E-07 |
| 205 | 3538.B20.GZ43_504848 | AB069994 | Macaca fascicularis testis cDNA clone:QtsA-10636, full insert sequence | 1.40E-07 |
| 206 | 3538.C01.GZ43_504545 | AF072375 | Pseudoalteromonas sp. S9 beta-hexosaminidase (chiP) gene, complete cds | 1.50E-04 |
| 207 | 3538.C02.GZ43_504561 | AJ011271 | Human immunodeficiency virus type 2 partial env gene, isolate b1286 | 7.10E-08 |
| 208 | 3538.D06.GZ43_504626 | AF100765 | Oryza sativa receptor-like kinase (8ARK1) gene, complete cds | 3.00E-06 |
| 209 | 3538.D09.GZ43_504674 | Z47784 | M.musculus mRNA expressed in islet cells (clone 58) | 3.40E-08 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 210 | 3538.D21.GZ43_504866 | AE006568 | Streptococcus pyogenes M1 GAS strain SF370, section 97 of 167 of the complete genome | 6.00E-07 |
| 211 | 3538.E15.GZ43_504771 | AB027966 | Schizosaccharomyces pombe gene for Hypothetical protein, partial cds, clone:TB89 | 2.30E-08 |
| 212 | 3538.F02.GZ43_504564 | AK001871 | Homo sapiens cDNA FLJ11009 fis, clone PLACE1003108 | 5.90E-09 |
| 213 | 3538.F08.GZ43_504660 | U39161 | Human phosphodiesterase (PDEA) gene, intron 8, 5' end | 7.40E-07 |
| 214 | 3553.K23.GZ43_506825 | Y12052 | Homo sapiens gene encoding guanine nucleotide-binding protein beta3 subunit, exon 5 | 3.00E-06 |
| 215 | 3553.K24.GZ43_506841 | AP001419 | Homo sapiens genomic DNA, chromosome 21q22.2, clone:PAC24K9, LB7T-ERG region, complete sequence | 1.00E-06 |
| 216 | 3553.L02.GZ43_506490 | X15028 | Chicken hsp90 gene for 90 kDa-heat shock protein 5'-end | 3.60E-05 |
| 217 | 3553.L04.GZ43_506522 | L34665 | Rattus norvegicus H+/K+-ATPase beta subunit (HKB) gene, exon 6 | 1.20E-09 |
| 218 | 3553.L21.GZ43_506794 | M87843 | Human transforming growth factor beta-2 gene, 5' end | 2.30E-05 |
| 219 | 3553.M12.GZ43_506651 | AB026592 | Limnoporus esakii mitochondrial gene for 16S ribosomal RNA, partial sequence | 1.10E-07 |
| 220 | 3553.M23.GZ43_506827 | AE006349 | Lactococcus lactis subsp. lactis IL1403 section 111 of 218 of the complete genome | 8.00E-07 |
| 221 | 3553.N01.GZ43_506476 | U80457 | Human transcription factor SIM2 short form mRNA, complete cds | 2.00E-06 |
| 222 | 3553.N02.GZ43_506492 | U81144 | Caenorhabditis elegans non-alpha nicotinic acetylcholine receptor subunit precursor (unc-29) gene, complete cds | 3.20E-07 |
| 223 | 3553.N04.GZ43_506524 | AE006296 | Lactococcus lactis subsp. lactis IL1403 section 58 of 218 of the complete genome | 2.00E-06 |
| 224 | 3553.N07.GZ43_506572 | AK026258 | Homo sapiens cDNA: FLJ22605 fis, clone HSI04743 | 1.00E-06 |
| 225 | 3553.N08.GZ43_506588 | U05822 | Human proto-oncogene BCL3 gene, exon 2 | 1.90E-14 |
| 226 | 3553.O07.GZ43_506573 | X97196 | D.melanogaster X gene | 4.00E-06 |
| 227 | 3553.O18.GZ43_506749 | AE001146 | Borrelia burgdorferi (section 32 of 70) of the complete genome | 1.60E-05 |
| 228 | 3553.O23.GZ43_506829 | X63509 | Mus musculus partial L1 gene, exons 2-4 | 6.00E-06 |
| 229 | 3553.P03.GZ43_506510 | M24901 | Rabbit pulmonary surfactant-associated protein (SP-B) mRNA, complete cds | 4.20E-07 |
| 230 | 3553.P05.GZ43_506542 | AF239156 | Homo sapiens peptide deformylase-like protein mRNA, complete cds | 1.00E-06 |
| 231 | 3553.P12.GZ43_506654 | AF283753 | Acipenser persicus isolate cw203 cytochrome b gene, partial cds; mitochondrial gene for mitochondrial product | 3.90E-07 |

**Table 8**

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 232 | 3553.P18.GZ43_506750 | AK021312 | Mus musculus 13 days embryo stomach cDNA, RIKEN full-length enriched library, clone:D530039A21, full insert sequence | 3.50E-08 |
| 233 | 3553.P21.GZ43_506798 | AB044136 | Homo sapiens genomic DNA, clone:#7 | 4.00E-06 |
| 234 | 3556.A03.GZ43_506879 | X61084 | C.griseus rhodopsin gene for opsin protein | 4.30E-05 |
| 235 | 3556.A06.GZ43_506927 | L46904 | Homo sapiens (subclone 4_c6 from P1 H22) DNA sequence | 1.20E-08 |
| 236 | 3556.B06.GZ43_506928 | AK002041 | Homo sapiens cDNA FLJ11179 fis, clone PLACE1007450 | 1.40E-07 |
| 237 | 3556.B09.GZ43_506976 | U88832 | Human groucho protein homolog (AES) gene, exons 2-7 and complete cds | 7.00E-07 |
| 238 | 3556.B10.GZ43_506992 | M11925 | Influenza A/chicken/Pennsylvania/8125/83 (H5N2) neuraminidase (NA) gene, complete cds | 5.00E-06 |
| 239 | 3556.B14.GZ43_507056 | Z80218 | Caenorhabditis elegans cosmid F52D4, complete sequence | 2.20E-05 |
| 240 | 3556.C13.GZ43_507041 | AF348512 | Mus musculus polyamine-modulated factor-1 gene, exons 2 through 5 and complete cds | 8.00E-06 |
| 241 | 3556.C15.GZ43_507073 | X82013 | S.cerevisiae mRNA for SUL1 | 3.00E-06 |
| 242 | 3556.C18.GZ43_507121 | Z23973 | H. sapiens (D7S660) DNA segment containing (CA) repeat; clone AFM277vd5; single read | 5.00E-06 |
| 243 | 3556.C24.GZ43_507217 | AE001381 | Plasmodium falciparum chromosome 2, section 18 of 73 of the complete sequence | 6.90E-07 |
| 244 | 3556.D15.GZ43_507074 | U48288 | Rattus norvegicus A-kinase anchoring protein AKAP 220 mRNA, complete cds | 5.50E-07 |
| 245 | 3556.D20.GZ43_507154 | AF092684 | Neochlamisus scabripennis haplotype 113 cytochrome oxidase I (COI) gene, mitochondrial gene encoding mitochondrial protein, partial cds | 4.00E-07 |
| 246 | 3556.D23.GZ43_507202 | X16416 | Human c-abl mRNA encoding p150 protein | 2.25E-04 |
| 247 | 3556.E13.GZ43_507043 | AL049948 | Homo sapiens mRNA; cDNA DKFZp564K0222 (from clone DKFZp564K0222) | 6.60E-08 |
| 248 | 3556.E24.GZ43_507219 | Z57634 | H.sapiens CpG island DNA genomic Mse1 fragment, clone 187e9, forward read cpg187e9.ft1a | 8.70E-07 |
| 249 | 3556.F10.GZ43_506996 | AF025409 | Homo sapiens zinc transporter 4 (ZNT4) mRNA, complete cds | 3.90E-34 |
| 250 | 3556.G15.GZ43_507077 | X15407 | Maize pseudo-Gpa2 pseudogene for glyceraldehyde-3-phosphate dehydrogenase subunit A | 3.40E-05 |
| 251 | 3556.H01.GZ43_506854 | AF269443 | Staphylococcus epidermidis strain SR1 clone step.1003h04 genomic sequence | 3.00E-06 |

**Table 8**

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 252 | 3556.H02.GZ43_506870 | U31465 | Kluyveromyces lactis telomerase RNA component (TER1) gene, complete sequence | 3.00E-06 |
| 253 | 3556.H12.GZ43_507030 | Z68886 | Human DNA sequence from cosmid L21F12, Huntington's Disease Region, chromosome 4p16.3 | 1.70E-07 |
| 254 | 3556.H20.GZ43_507158 | AB034628 | Equus caballus microsatellite TKY319, TKY320 DNA | 1.70E-07 |
| 255 | 3556.I02.GZ43_506871 | AL390767 | Human DNA sequence from clone RP1-68P15 on chromosome 11p13-14.2 Contains GSSs and ESTs. Contains part of a novel gene, complete sequence [Homo sapiens] | 2.00E-06 |
| 256 | 3556.I14.GZ43_507063 | U34042 | Mus musculus mammalian tolloid-like protein mRNA, complete cds | 1.50E-05 |
| 257 | 3556.J05.GZ43_506920 | U31465 | Kluyveromyces lactis telomerase RNA component (TER1) gene, complete sequence | 2.00E-06 |
| 258 | 3556.J07.GZ43_506952 | AL359621 | Homo sapiens mRNA; cDNA DKFZp434M1631 (from clone DKFZp434M1631) | 2.00E-06 |
| 259 | 3556.J14.GZ43_507064 | M81830 | Human somatostatin receptor isoform 2 (SSTR2) gene, complete cds | 1.00E-06 |
| 260 | 3556.J16.GZ43_507096 | Y15484 | Canis familiaris gene encoding retinal guanylate cyclase E | 2.90E-08 |
| 261 | 3556.K04.GZ43_506905 | U88832 | Human groucho protein homolog (AES) gene, exons 2-7 and complete cds | 8.00E-07 |
| 262 | 3556.K12.GZ43_507033 | AP001419 | Homo sapiens genomic DNA, chromosome 21q22.2, clone:PAC24K9, LB7T-ERG region, complete sequence | 1.00E-06 |
| 263 | 3556.K13.GZ43_507049 | AK023589 | Homo sapiens cDNA FLJ13527 fis, clone PLACE1006076 | 2.00E-06 |
| 264 | 3556.K17.GZ43_507113 | X71634 | D.bifasciata P-Transposon | 3.00E-06 |
| 265 | 3556.L08.GZ43_506970 | X02367 | Glaucoma chattoni rDNA 3' NTS | 8.20E-08 |
| 266 | 3556.L09.GZ43_506986 | AF154329 | Pisum sativum MAP kinase PsMAPK2 (Mapk2) mRNA, complete cds | 4.10E-07 |
| 267 | 3556.L16.GZ43_507098 | AB041791 | Homo sapiens HSPDE10A gene for phosphodiesterase 10A1 (PDE10A1), exon 17 | 3.10E-08 |
| 268 | 3556.L23.GZ43_507210 | M23720 | Rat carboxypeptidase (CA2) gene, exon 10 | 5.00E-06 |
| 269 | 3556.M02.GZ43_506875 | U91963 | Human tolloid-like protein (TLL) mRNA, complete cds | 1.40E-05 |
| 270 | 3556.M11.GZ43_507019 | X16353 | R.rickettsii ompB gene for outer membrane protein B | 7.60E-05 |
| 271 | 3556.M23.GZ43_507211 | X93496 | H.sapiens TRAP gene, 5' flanking region | 5.60E-23 |
| 272 | 3556.N02.GZ43_506876 | U26458 | Snakehead retrovirus (SnRV), complete genome | 3.20E-05 |
| 273 | 3556.N04.GZ43_506908 | L39064 | Homo sapiens interleukin 9 receptor precursor (IL9R) gene, complete cds | 4.00E-09 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 274 | 3556.N05.GZ43_506924 | M63437 | Chicken KLG gene, complete cds | 2.00E-06 |
| 275 | 3556.N06.GZ43_506940 | AF327424 | Arabidopsis thaliana unknown protein (T14P1.19/At2g45010) mRNA, partial cds | 2.00E-07 |
| 276 | 3556.N21.GZ43_507180 | AB022157 | Mus musculus Cctd gene for chaperonin containing TCP-1 delta subunit, complete cds | 4.00E-06 |
| 277 | 3556.O08.GZ43_506973 | X00171 | Vibrio cholera toxin (ctx) operon DNA sequence from strain 2125 | 7.00E-06 |
| 278 | 3556.O13.GZ43_507053 | U41106 | Caenorhabditis elegans cosmid W06A11 | 1.20E-05 |
| 279 | 3556.P07.GZ43_506958 | M15085 | T.brucei expressed copy of the ILTat 1.3 variable surface glycoprotein gene, 5' flank | 1.10E-07 |
| 280 | 3559.A04.GZ43_507279 | AE006824 | Sulfolobus solfataricus section 183 of 272 of the complete genome | 4.70E-05 |
| 281 | 3559.A20.GZ43_507535 | X71787 | A.thaliana AAP2 mRNA for amino acid permease | 2.00E-06 |
| 282 | 3559.A24.GZ43_507599 | X56494 | H.sapiens M gene for M1-type and M2-type pyruvate kinase | 1.80E-05 |
| 283 | 3559.B04.GZ43_507280 | AJ251550 | Homo sapiens partial AK155 gene for AK155 protein, exons 1-3 and joined CDS | 2.50E-05 |
| 284 | 3559.B06.GZ43_507312 | AF077344 | Homo sapiens cartilage-derived C-type lectin (CLECSF1) gene, exons 1 and 2 | 5.80E-05 |
| 285 | 3559.B08.GZ43_507344 | D50552 | Xenopus laevis xSox12 mRNA for XSOX12, complete cds | 4.00E-07 |
| 286 | 3559.B10.GZ43_507376 | L76259 | Homo sapiens PTS gene, complete cds | 9.00E-06 |
| 287 | 3559.B18.GZ43_507504 | M29109 | D.discoideum actin M6 gene, 5' flank | 3.40E-07 |
| 288 | 3559.C06.GZ43_507313 | X99910 | C.carpio mRNA transcription factor, ovx1 | 1.60E-05 |
| 289 | 3559.D21.GZ43_507554 | AK022877 | Homo sapiens cDNA FLJ12815 fis, clone NT2RP2002546 | 2.00E-06 |
| 290 | 3559.E06.GZ43_507315 | U97408 | Caenorhabditis elegans cosmid F48A9 | 3.00E-06 |
| 291 | 3559.E09.GZ43_507363 | L40489 | Ureaplasma urealyticum UreA (ureA), UreB (ureB), UreC (ureC), UreE (ureE), UreF (ureF), and UreG (ureG) genes, complete cds; UreD (ureD) gene, partial cds; and unknown gene | 3.00E-07 |
| 292 | 3559.E20.GZ43_507539 | AF113521 | Zea mays putative transcription factor mRNA sequence | 8.20E-08 |
| 293 | 3559.F07.GZ43_507332 | AF109377 | Mus musculus ldlBp (LDLB) mRNA, complete cds | 4.30E-05 |
| 294 | 3559.F17.GZ43_507492 | U11292 | Human Ki nuclear autoantigen mRNA, complete cds | 6.40E-07 |
| 295 | 3559.H09.GZ43_507366 | X13414 | Murine I gene for MHC class II(Ia) associated invariant chain | 9.00E-06 |
| 296 | 3559.H22.GZ43_507574 | U61402 | Streptococcus thermophilus GalR (galR), galactokinase (galK) and gal-1-P uridylyltransferase (galT) genes, complete cds | 1.00E-06 |
| 297 | 3559.H24.GZ43_507606 | U67594 | Methanococcus jannaschii section 136 of 150 of the complete genome | 3.40E-05 |
| 298 | 3559.I05.GZ43_507303 | X97289 | S.salar genes encoding alpha-globin and beta-globin, clone 6 | 7.00E-06 |

## Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 299 | 3559.J04.GZ43_507288 | L10709 | Human constitutive endothelial nitric oxide synthase gene, exons 25 and 26 and complete cds | 8.90E-12 |
| 300 | 3559.J20.GZ43_507544 | U67559 | Methanococcus jannaschii section 101 of 150 of the complete genome | 5.70E-05 |
| 301 | 3559.K16.GZ43_507481 | Z48955 | D.virginiana partial LINE-1 repetitive DNA and putative RT | 2.40E-08 |
| 302 | 3559.K17.GZ43_507497 | AC004497 | Homo sapiens chromosome 21, P1 clone LBNL#6 (LBNL H10), complete sequence | 4.00E-06 |
| 303 | 3559.L01.GZ43_507242 | X58774 | Herpesvirus saimiri sRNA1, sRNA2, sRNA3 and sRNA4 genes for small viral RNAs | 1.00E-06 |
| 304 | 3559.L14.GZ43_507450 | X67774 | C.upsaliensis (LMG 8854) 23S rRNA gene | 1.30E-05 |
| 305 | 3559.L19.GZ43_507530 | Z57634 | H.sapiens CpG island DNA genomic Msel fragment, clone 187e9, forward read cpg187e9.ft1a | 7.70E-07 |
| 306 | 3559.M02.GZ43_50725 9 | AF042834 | Homo sapiens phosphodiesterase delta subunit gene, exons 2, 3 and 4 | 1.30E-05 |
| 307 | 3559.M09.GZ43_50737 1 | U07628 | Caenorhabditis elegans N2 APX-1 (apx-1) mRNA, complete cds | 2.00E-06 |
| 308 | 3559.N05.GZ43_507308 | Z24259 | H. sapiens (D19S417) DNA segment containing (CA) repeat; clone AFM304zg1; single read | 3.70E-07 |
| 309 | 3559.N18.GZ43_507516 | S75829 | {dinucleotide repeats, microsatellite marker} [Dryobalanops lanceolata, Genomic, 230 nt] | 1.90E-07 |
| 310 | 3559.N21.GZ43_507564 | AL353948 | Homo sapiens mRNA; cDNA DKFZp761P0114 (from clone DKFZp761P0114) | 5.30E-07 |
| 311 | 3559.O01.GZ43_507245 | AL110269 | Homo sapiens mRNA; cDNA DKFZp564A122 (from clone DKFZp564A122); partial cds | 1.60E-17 |
| 312 | 3559.O05.GZ43_507309 | Y08695 | Clostridium tertium nanH gene | 7.40E-07 |
| 313 | 3559.O07.GZ43_507341 | AJ249489 | Xenopus laevis partial mRNA for putative olfactory receptor (xb6 gene) | 5.40E-07 |
| 314 | 3559.O20.GZ43_507549 | X02886 | Human gene for T-cell receptor alpha chain J region | 2.00E-06 |
| 315 | 3559.P10.GZ43_507390 | X66030 | Homo sapiens partial ufo gene encoding tyrosine kinase receptor | 4.90E-07 |
| 316 | 3559.P15.GZ43_507470 | Z16777 | H. sapiens (D2S139) DNA segment containing (CA) repeat; clone AFM177xh4; single read | 4.00E-06 |
| 317 | 3559.P18.GZ43_507518 | AJ228072 | Nicotiana benthamiana DNA for Tnt1 retrotransposable element, isolate ben15 | 2.80E-07 |
| 318 | 3559.P24.GZ43_507614 | U32372 | Rattus norvegicus tyrosine-ester sulfotransferase mRNA, complete cds | 4.90E-07 |
| 319 | 3562.A01.GZ43_507615 | AE000496 | Escherichia coli K12 MG1655 section 386 of 400 of the complete genome | 1.56E-04 |
| 320 | 3562.A15.GZ43_507839 | AF068289 | Homo sapiens HDCMD34P mRNA, complete cds | 6.60E-11 |

**Table 8**

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 321 | 3562.B22.GZ43_507952 | AK014534 | Mus musculus 0 day neonate skin cDNA, RIKEN full-length enriched library, clone:4631424J17, full insert sequence | 1.10E-07 |
| 322 | 3562.C23.GZ43_507969 | L01787 | Ascaris suum phosphoenolpyruvate carboxykinase (PEPCK) gene, complete cds | 3.70E-07 |
| 323 | 3562.D10.GZ43_507762 | X54061 | D. melanogaster mRNA coding for a 205K microtubule-associated protein (MAP) | 6.60E-07 |
| 324 | 3562.E01.GZ43_507619 | M29812 | Homo sapiens Ig H-chain V71-4 (IGH@) gene, partial cds | 1.50E-05 |
| 325 | 3562.E03.GZ43_507651 | X03729 | Vaccinia virus late gene cluster from central portion of genome containing the L65 gene locus | 2.63E-04 |
| 326 | 3562.E12.GZ43_507795 | M29694 | B.licheniformis RNA polymerase sigma-30 factor (spo0H) gene, complete cds | 1.60E-05 |
| 327 | 3562.F19.GZ43_507908 | AE006216 | Pasteurella multocida PM70 section 183 of 204 of the complete genome | 2.40E-05 |
| 328 | 3562.F20.GZ43_507924 | M91004 | Rabbit endothelial leukocyte adhesion molecule I (ELAM1), complete cds | 2.00E-06 |
| 329 | 3562.G13.GZ43_507813 | X69818 | E.muelleri COLF1 gene for extracellular matrix protein | 1.00E-06 |
| 330 | 3562.G19.GZ43_507909 | AK019034 | Mus musculus 10 day old male pancreas cDNA, RIKEN full-length enriched library, clone:1810049K24, full insert sequence | 1.00E-05 |
| 331 | 3562.H11.GZ43_507782 | AF206598 | Algyroides fitzingeri 12S ribosomal RNA gene, partial sequence; tRNA-Val gene, complete sequence; and 16S ribosomal RNA gene, partial sequence; mitochondrial genes for mitochondrial products | 1.40E-07 |
| 332 | 3562.H12.GZ43_507798 | AK002041 | Homo sapiens cDNA FLJ11179 fis, clone PLACE1007450 | 5.30E-07 |
| 333 | 3562.I01.GZ43_507623 | S39048 | knob associated histidine-rich protein KAHRP {5'region} [Plasmodium falciparum, Genomic, 2215 nt] | 2.00E-06 |
| 334 | 3562.I02.GZ43_507639 | AF129501 | Buchnera aphidicola natural-host Diuraphis noxia acetohydroxy acid synthase large subunit (ilvI) and acetohydroxy acid synthase small subunit (ilvH) genes, complete cds; and unknown genes | 1.60E-07 |
| 335 | 3562.I13.GZ43_507815 | M26049 | Yeast (S.cerevisiae) RAD9 protein (required for cell cycle arrest during DNA repair) gene, complete cds | 4.00E-06 |
| 336 | 3562.I15.GZ43_507847 | AF310880 | Barbatula barbatula microsatellite Bbar5 sequence | 1.60E-07 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 337 | 3562.J09.GZ43_507752 | AF236642 | Calothrix parietina clone 102-2A 16S-23S internal transcribed spacer, complete sequence; and tRNA-Ile and tRNA-Ala genes, complete sequence | 3.30E-07 |
| 338 | 3562.J13.GZ43_507816 | AC010728 | Homo sapiens BAC clone RP11-258E22 from Y, complete sequence | 1.30E-05 |
| 339 | 3562.K04.GZ43_507673 | S79777 | {specific DNA probe for Plasmodium vivax pARC 1153} [Plasmodium vivax, host=human, Genomic, 665 nt] | 5.40E-07 |
| 340 | 3562.K08.GZ43_507737 | AJ403240 | M.musculus DNA for vimentin-binding fragment VimE8 | 2.00E-06 |
| 341 | 3562.L12.GZ43_507802 | AE007840 | Clostridium acetobutylicum ATCC824 section 328 of 356 of the complete genome | 5.80E-07 |
| 342 | 3562.N24.GZ43_507996 | AF255609 | Homo sapiens high mobility group protein HMG1 gene, exons 1 and 2, partial cds | 2.00E-07 |
| 343 | 3562.O11.GZ43_507789 | M15027 | Human myelin proteolipid protein gene, exon 2 | 1.00E-06 |
| 344 | 3562.O18.GZ43_507901 | AL050208 | Homo sapiens mRNA; cDNA DKFZp586F2323 (from clone DKFZp586F2323) | 2.40E-07 |
| 345 | 3562.O20.GZ43_507933 | AY020756 | Oryza sativa microsatellite MRG3081 containing (TA)X13, genomic sequence | 4.90E-08 |
| 346 | 3562.P21.GZ43_507950 | AF036318 | Skeletonema costatum cyclin (CYCL) gene, partial cds | 7.20E-07 |
| 347 | 3562.P23.GZ43_507982 | AF126719 | Plasmodium falciparum cAMP-dependent protein kinase (pka) gene, complete cds | 3.00E-06 |
| 348 | 3565.A23.GZ43_508351 | AL122065 | Homo sapiens mRNA; cDNA DKFZp434N011 (from clone DKFZp434N011) | 1.50E-07 |
| 349 | 3565.B05.GZ43_508064 | AF163325 | Trichoderma harzianum mitochondrial plasmid pThr1, complete plasmid sequence | 1.50E-07 |
| 350 | 3565.B13.GZ43_508192 | X62689 | T.retusa DNA for brachiopod cubitus-interruptus dominant (ciD) homologue | 9.00E-06 |
| 351 | 3565.B14.GZ43_508208 | M29929 | Human insulin receptor (allele 1) gene, exons 14, 15, 16 and 17 | 4.30E-12 |
| 352 | 3565.C04.GZ43_508049 | AE006183 | Pasteurella multocida PM70 section 150 of 204 of the complete genome | 2.00E-06 |
| 353 | 3565.C06.GZ43_508081 | S79836 | SCPx/SCP2=sterol carrier protein x/sterol carrier protein 2 {promoter} [human, Genomic, 3575 nt] | 3.00E-06 |
| 354 | 3565.C17.GZ43_508257 | L13937 | Bovine phospholipase C mRNA, complete cds | 3.00E-07 |
| 355 | 3565.D14.GZ43_508210 | M37818 | Human keratin (psi-K-alpha) pseudogene, exons 4,5,6,7 and 8, and keratin (psi-K-beta) pseudogene, complete cds | 3.50E-08 |
| 356 | 3565.D17.GZ43_508258 | Z19005 | C.pasteurianum gene for ferredoxin | 1.00E-06 |

**Table 8**

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 357 | 3565.D19.GZ43_508290 | AE007758 | Clostridium acetobutylicum ATCC824 section 246 of 356 of the complete genome | 3.00E-06 |
| 358 | 3565.E16.GZ43_508243 | L42813 | Protopterus dolloi complete mitochondrial genome | 2.49E-04 |
| 359 | 3565.G07.GZ43_508101 | U97500 | Homo sapiens butyrophilin (BT3.3) gene, exons 1-4 | 1.30E-05 |
| 360 | 3565.G09.GZ43_508133 | M95098 | Bos taurus lysozyme gene (cow 2), complete cds | 1.26E-04 |
| 361 | 3565.G22.GZ43_508341 | AJ400873 | Homo sapiens partial GPLD1 gene for glycosylphosphatidylinositol phospholipase D, exons 15-20 | 1.40E-09 |
| 362 | 3565.H06.GZ43_508086 | U67465 | Methanococcus jannaschii section 7 of 150 of the complete genome | 6.10E-07 |
| 363 | 3565.H10.GZ43_508150 | M15350 | Bacillus sp. strain 170 beta-lactamase gene, complete cds | 5.70E-08 |
| 364 | 3565.H11.GZ43_508166 | AB044878 | Equus caballus DNA, microsatellite TKY378 | 3.20E-09 |
| 365 | 3565.H15.GZ43_508230 | AL122122 | Homo sapiens mRNA; cDNA DKFZp434L098 (from clone DKFZp434L098) | 5.00E-06 |
| 366 | 3565.H23.GZ43_508358 | J05492 | E.coli cytochrome O ubiquinol oxidase (cyoA, cyoB, cyoC, cyoD and cyoE genes, complete cds | 1.00E-06 |
| 367 | 3565.H24.GZ43_508374 | AE001417 | Plasmodium falciparum chromosome 2, section 54 of 73 of the complete sequence | 1.70E-10 |
| 368 | 3565.K15.GZ43_508233 | AB062985 | Macaca fascicularis brain cDNA clone:QmoA-10670, full insert sequence | 6.90E-105 |
| 369 | 3565.L22.GZ43_508346 | L81801 | Homo sapiens (subclone 1_a2 from P1 H31) DNA sequence, complete sequence | 1.30E-05 |
| 370 | 3565.M15.GZ43_508235 | X08038 | Methanobacterium thermoautotrophicum rpoT, rpoU, rpoV and rpoX genes for RNA polymerase subunits A, B', B'' and C | 1.10E-05 |
| 371 | 3565.M20.GZ43_508315 | Z93381 | Caenorhabditis elegans cosmid F28G4, complete sequence | 1.20E-05 |
| 372 | 3565.N12.GZ43_508188 | M21573 | Salmon (S.salar) growth hormone gene, complete cds | 5.70E-05 |
| 373 | 3565.N13.GZ43_508204 | AK001163 | Homo sapiens cDNA FLJ10301 fis, clone NT2RM2000032 | 5.20E-08 |
| 374 | 3565.N19.GZ43_508300 | AF321321 | Homo sapiens dopamine transporter (SLC6A3) gene, exon 15 and complete cds | 2.00E-06 |
| 375 | 3565.O02.GZ43_508029 | X59773 | Pisum sativum mRNA for P protein, a part of glycine cleavage complex | 1.30E-05 |
| 376 | 3565.O03.GZ43_508045 | Z27113 | H.Sapiens gene for RNA polymerase II subunit 14.4 kD | 2.00E-15 |
| 377 | 3565.O07.GZ43_508109 | X96607 | M.musculus IgH 3' alpha enhancer DNA | 6.40E-05 |
| 378 | 3565.O15.GZ43_508237 | Z35484 | Thermoanaerobacter sp. ATCC53627 cgtA gene | 3.00E-06 |

**Table 8**

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 379 | 3565.P03.GZ43_508046 | U11292 | Human Ki nuclear autoantigen mRNA, complete cds | 6.40E-07 |
| 380 | 3565.P09.GZ43_508142 | X56261 | Yeast PPH1 gene for protein phosphatase 2A | 1.00E-06 |
| 381 | 3565.P22.GZ43_508350 | AE007790 | Clostridium acetobutylicum ATCC824 section 278 of 356 of the complete genome | 3.00E-06 |
| 382 | 3565.P24.GZ43_508382 | X61146 | N.tabacum NTP303 pollen specific mRNA | 2.70E-05 |
| 383 | 3568.A10.GZ43_508545 | U46925 | Arabidopsis thaliana GTP-binding protein ATGB2 mRNA, complete cds | 3.00E-06 |
| 384 | 3568.B02.GZ43_508418 | U83640 | Mus caroli Sp100 gene, exons 3 and 4 | 1.90E-08 |
| 385 | 3568.B05.GZ43_508466 | BC008293 | Homo sapiens, Similar to RIKEN cDNA A430101B06 gene, clone MGC:13017 IMAGE:3537789, mRNA, complete cds | 3.20E-16 |
| 386 | 3568.C22.GZ43_508739 | AF280797 | Homo sapiens NPC-related protein NAG73 mRNA, complete cds | 1.00E-06 |
| 387 | 3568.D23.GZ43_508756 | AK022922 | Homo sapiens cDNA FLJ12860 fis, clone NT2RP2003559 | 8.00E-06 |
| 388 | 3568.E17.GZ43_508661 | AF068294 | Homo sapiens HDCMB45P mRNA, partial cds | 5.30E-09 |
| 389 | 3568.E20.GZ43_508709 | AE006417 | Lactococcus lactis subsp. lactis IL1403 section 179 of 218 of the complete genome | 1.10E-05 |
| 390 | 3568.F06.GZ43_508486 | U52198 | Vibrio anguillarum flagellin E (flaE), flagellin D (flaD), and flagellin B (flaB) genes, complete cds, and (flaG) gene, partial cds | 2.20E-05 |
| 391 | 3568.F07.GZ43_508502 | Z23599 | H. sapiens (D13S263) DNA segment containing (CA) repeat; clone AFM210yg11; single read | 1.90E-08 |
| 392 | 3568.F11.GZ43_508566 | AE007525 | Clostridium acetobutylicum ATCC824 section 13 of 356 of the complete genome | 4.20E-07 |
| 393 | 3568.F12.GZ43_508582 | D50416 | Mouse mRNA for AREC3, complete cds | 1.90E-05 |
| 394 | 3568.F22.GZ43_508742 | AF025900 | Histrionicus histrionicus CA dinucleotide repeat locus Hhimicro1 | 7.80E-07 |
| 395 | 3568.G10.GZ43_508551 | U66074 | Tritrichomonas foetus putative superoxide dismutase 2 (SOD2) gene, complete cds | 9.70E-07 |
| 396 | 3568.G12.GZ43_508583 | AB062941 | Macaca fascicularis brain cDNA clone:QflA-14927, full insert sequence | 9.50E-47 |
| 397 | 3568.G24.GZ43_508775 | L27221 | Giardia intestinalis pyruvate:flavodoxin oxidoreductase and flanking genes | 3.20E-05 |
| 398 | 3568.H20.GZ43_508712 | X75887 | B.taurus Brevican mRNA | 4.70E-05 |
| 399 | 3568.J10.GZ43_508554 | AF194829 | Tetragonia tetragonioides NADH dehydrogenase (ndhF) gene, partial cds; chloroplast gene for chloroplast product | 2.00E-06 |
| 400 | 3568.J22.GZ43_508746 | Y11031 | C.coli pldA gene | 1.00E-06 |
| 401 | 3568.K01.GZ43_508411 | AL137751 | Homo sapiens mRNA; cDNA DKFZp434I0812 (from clone DKFZp434I0812); partial cds | 3.00E-06 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 402 | 3568.K04.GZ43_508459 | Z82295 | R.prowazekii genomic DNA fragment (clone A153F) | 7.20E-08 |
| 403 | 3568.L04.GZ43_508460 | AL050105 | Homo sapiens mRNA; cDNA DKFZp586H0519 (from clone DKFZp586H0519); partial cds | 1.00E-05 |
| 404 | 3568.M03.GZ43_508445 | L76259 | Homo sapiens PTS gene, complete cds | 8.00E-06 |
| 405 | 3568.M13.GZ43_508605 | X61218 | M.musculus cervicolor (strain CRP) Tcp-1 gene for t-complex polypeptide 1, exons 8-10 | 3.10E-09 |
| 406 | 3568.N11.GZ43_508574 | AL079296 | Homo sapiens mRNA full length insert cDNA clone EUROIMAGE 609395 | 2.00E-06 |
| 407 | 3568.O17.GZ43_508671 | AF078848 | Homo sapiens BUP mRNA, complete cds | 9.50E-09 |
| 408 | 3568.P04.GZ43_508464 | AB041548 | Mus musculus brain cDNA, clone MNCb-3816, similar to AF171875 g1-related zinc finger protein (Mus musculus) | 5.00E-06 |
| 409 | 3568.P18.GZ43_508688 | AL358951 | Human DNA sequence from clone RP3-456L16 on chromosome 6, complete sequence [Homo sapiens] | 3.00E-07 |
| 410 | 3568.P19.GZ43_508704 | U43542 | Nicotiana tabacum diphenol oxidase mRNA, complete cds | 2.00E-06 |
| 411 | 3571.A04.GZ43_508833 | AF017116 | Homo sapiens type-2 phosphatidic acid phosphohydrolase (PAP2) mRNA, complete cds | 2.40E-07 |
| 412 | 3571.A07.GZ43_508881 | L81867 | Homo sapiens (subclone 1_a8 from P1 H54) DNA sequence, complete sequence | 9.00E-06 |
| 413 | 3571.A08.GZ43_508897 | X85041 | H.sapiens PE5L gene ALU repeat region | 2.00E-06 |
| 414 | 3571.A11.GZ43_508945 | U19361 | Petromyzon marinus neurofilament subunit NF-180 mRNA, complete cds | 4.70E-08 |
| 415 | 3571.A14.GZ43_508993 | AL022342 | Human DNA sequence from clone RP1-29M10 on chromosome 20, complete sequence [Homo sapiens] | 7.00E-05 |
| 416 | 3571.A22.GZ43_509121 | U09448 | Vaucheria bursata protein synthesis elongation factor Tu (tufA) gene, chloroplast gene encoding chloroplast protein, partial cds | 7.20E-07 |
| 417 | 3571.B13.GZ43_508978 | AE002555 | Neisseria meningitidis serogroup B strain MC58 section 197 of 206 of the complete genome | 4.40E-05 |
| 418 | 3571.B22.GZ43_509122 | AE002555 | Neisseria meningitidis serogroup B strain MC58 section 197 of 206 of the complete genome | 4.50E-05 |
| 419 | 3571.C08.GZ43_508899 | AJ010154 | Saguinus oedipus msp-E1 gene | 1.10E-17 |
| 420 | 3571.D04.GZ43_508836 | AF125460 | Caenorhabditis elegans cosmid Y9D1A | 3.60E-07 |
| 421 | 3571.D07.GZ43_508884 | U51654 | Barbus barbus x Barbus meridionalis microsatellite clone no.37 | 8.72E-02 |
| 422 | 3571.E02.GZ43_508805 | AF329081 | Bos taurus AMP-activated protein kinase gamma-1 (PRKAG1) gene, partial cds | 4.40E-33 |
| 423 | 3571.E10.GZ43_508933 | M96068 | Madagascar periwinkle hydroxymethylglutaryl-CoA reductase (HMGR) mRNA, complete cds | 3.30E-08 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 424 | 3571.E16.GZ43_509029 | AE006429 | Lactococcus lactis subsp. lactis IL1403 section 191 of 218 of the complete genome | 1.30E-05 |
| 425 | 3571.F06.GZ43_508870 | AL137296 | Homo sapiens mRNA; cDNA DKFZp434M0416 (from clone DKFZp434M0416) | 4.40E-07 |
| 426 | 3571.F16.GZ43_509030 | M58478 | Human cystic fibrosis transmembrane conductance regulator gene, 5' end | 6.30E-05 |
| 427 | 3571.F23.GZ43_509142 | AF038397 | Mus musculus glutaminase (Gls) gene, partial 3' sequence | 4.70E-08 |
| 428 | 3571.G22.GZ43_509127 | M80596 | Saccharomyces cerevisiae VAC1 gene (required for vacuole inheritance and vacuole protein sorting), complete cds | 7.00E-06 |
| 429 | 3571.G24.GZ43_509159 | Z75330 | H.sapiens mRNA for nuclear protein SA-1 | 1.00E-46 |
| 430 | 3571.H01.GZ43_508792 | U71144 | Influenza A virus H3N2 A/Akita/1/94 nucleoprotein (NP) gene, complete cds | 1.90E-05 |
| 431 | 3571.H10.GZ43_508936 | AF038564 | Homo sapiens atrophin-1 interacting protein 4 (AIP4) mRNA, partial cds | 6.60E-53 |
| 432 | 3571.H12.GZ43_508968 | K00131 | mouse b2 repeat sequence from clone mm61 | 3.00E-08 |
| 433 | 3571.H16.GZ43_509032 | AF179564 | Homo sapiens GTF2I-like sequence within duplicated segment of Williams syndrome region | 1.20E-23 |
| 434 | 3571.H18.GZ43_509064 | AE000331 | Escherichia coli K12 MG1655 section 221 of 400 of the complete genome | 1.45E-04 |
| 435 | 3571.I11.GZ43_508953 | U20661 | Dictyostelium discoideum unknown internal repeat protein gene, complete cds, and unknown orf1, orf2 and orf3 genes, partial cds | 9.00E-06 |
| 436 | 3571.J07.GZ43_508890 | M58478 | Human cystic fibrosis transmembrane conductance regulator gene, 5' end | 6.40E-05 |
| 437 | 3571.J08.GZ43_508906 | AK021312 | Mus musculus 13 days embryo stomach cDNA, RIKEN full-length enriched library, clone:D530039A21, full insert sequence | 3.60E-08 |
| 438 | 3571.J09.GZ43_508922 | X66483 | D.discoideum gp80 gene | 8.90E-07 |
| 439 | 3571.J14.GZ43_509002 | L77119 | Methanococcus jannaschii small extra-chromosomal element, complete sequence.~ | 1.40E-05 |
| 440 | 3571.L01.GZ43_508796 | AK005500 | Mus musculus adult female placenta cDNA, RIKEN full-length enriched library, clone:1600019O04, full insert sequence | 6.00E-06 |
| 441 | 3571.M17.GZ43_509053 | AF085681 | Mus musculus tubby like protein 1 (Tulp1) mRNA, complete cds | 5.00E-06 |
| 442 | 3571.M19.GZ43_509085 | D10487 | B.thermoglucosidasius gene for oligo-1,6-glucosidase | 9.00E-06 |
| 443 | 3571.M24.GZ43_509165 | M97680 | Bluetongue virus type 2 genomic RNA sequence | 2.00E-06 |
| 444 | 3571.N09.GZ43_508926 | X86100 | R.norvegicus BSP gene | 3.40E-07 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 445 | 3571.N14.GZ43_509006 | D32007 | Mouse mRNA for a homlogue of human CBFA2T1(Mtg8a), complete cds | 1.20E-08 |
| 446 | 3571.N17.GZ43_509054 | Z68755 | Human DNA sequence from cosmid L118D5, Huntington's Disease Region, chromosome 4p16.3 | 1.70E-10 |
| 447 | 3571.N22.GZ43_509134 | D00326 | Porcine rotavirus (strain Gottfried), VP6 gene, complete cds | 1.00E-06 |
| 448 | 3571.O08.GZ43_508911 | X66483 | D.discoideum gp80 gene | 8.20E-07 |
| 449 | 3574.A20.GZ43_509473 | AJ271814 | Drosophila melanogaster mRNA for meso18E protein | 1.70E-07 |
| 450 | 3574.B01.GZ43_509170 | U93261 | Homo sapiens DESP4P1 pseudogene sequence | 1.00E-06 |
| 451 | 3574.B04.GZ43_509218 | Y08207 | C.elaphus mitochondrial tRNA-Thr, tRNA-Pro and tRNA-Phe genes | 1.40E-14 |
| 452 | 3574.B10.GZ43_509314 | AL161991 | Homo sapiens mRNA; cDNA DKFZp761C169 (from clone DKFZp761C169); partial cds | 3.00E-06 |
| 453 | 3574.B14.GZ43_509378 | D79208 | Apis mellifera mRNA for alpha-glucosidase, complete cds | 7.00E-06 |
| 454 | 3574.B24.GZ43_509538 | AE007758 | Clostridium acetobutylicum ATCC824 section 246 of 356 of the complete genome | 3.00E-06 |
| 455 | 3574.C09.GZ43_509299 | AF057708 | Populus balsamifera subsp. trichocarpa PTD protein (PTD) gene, complete cds | 2.40E-07 |
| 456 | 3574.C10.GZ43_509315 | AE005602 | Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 221 of 290 | 9.70E-05 |
| 457 | 3574.C12.GZ43_509347 | AJ223633 | Enterococcus faecium genes encoding enterocin L50A and enterocin L50B plus 5' and 3' flanking regions | 9.50E-07 |
| 458 | 3574.C14.GZ43_509379 | X99710 | L.lactis ORF, genes homologous to vsf-1 and pepF2 and gene encoding protein homologous to methyltransferase | 4.00E-06 |
| 459 | 3574.C16.GZ43_509411 | AF092920 | Chlorohydra viridissima head-activator binding protein precursor (HAB) mRNA, complete cds | 3.00E-07 |
| 460 | 3574.C23.GZ43_509523 | AB047856 | Oryza sativa Ub-CEP52-2 gene for ubiquitin fused to ribosomal protein L40, complete cds | 5.00E-08 |
| 461 | 3574.D02.GZ43_509188 | AB060225 | Macaca fascicularis brain cDNA clone:QflA-14955, full insert sequence | 5.70E-07 |
| 462 | 3574.D12.GZ43_509348 | M58478 | Human cystic fibrosis transmembrane conductance regulator gene, 5' end | 6.00E-05 |
| 463 | 3574.E02.GZ43_509189 | L37347 | Human integral membrane protein (Nramp2) mRNA, partial | 2.00E-06 |
| 464 | 3574.E03.GZ43_509205 | X05817 | Bovine papillomavirus type 4 (BPV-4) genome | 6.00E-06 |
| 465 | 3574.E14.GZ43_509381 | U67507 | Methanococcus jannaschii section 49 of 150 of the complete genome | 3.40E-05 |
| 466 | 3574.F10.GZ43_509318 | M24376 | Mouse zinc finger protein (krox-20) gene, exon 1 | 3.80E-08 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 467 | 3574.F18.GZ43_509446 | AF184170 | Sparus aurata elongation factor 1-alpha (EF1-alpha) mRNA, complete cds | 3.40E-07 |
| 468 | 3574.F23.GZ43_509526 | Z29486 | R.norvegicus (Sprague Dawley) mRNA for AMP-activated protein kinase | 9.00E-06 |
| 469 | 3574.G07.GZ43_509271 | AF064079 | Plasmodium gallinaceum endochitinase precursor, mRNA, complete cds | 1.40E-07 |
| 470 | 3574.G11.GZ43_509335 | AF032872 | Rattus norvegicus potassium channel regulatory protein KChAP mRNA, complete cds | 7.40E-07 |
| 471 | 3574.H07.GZ43_509272 | J04718 | Human proliferating cell nuclear antigen (PCNA) gene, complete cds | 3.10E-07 |
| 472 | 3574.I02.GZ43_509193 | AF200361 | Rattus norvegicus cytochrome P450 4F1 (Cyp4F1) gene, complete cds | 1.50E-05 |
| 473 | 3574.I07.GZ43_509273 | M29688 | S.cerevisiae PMS1 gene encoding DNA mismatch repair protein, complete cds | 1.20E-08 |
| 474 | 3574.J11.GZ43_509338 | Z24104 | H. sapiens (D12S338) DNA segment containing (CA) repeat; clone AFM291wd9; single read | 3.20E-07 |
| 475 | 3574.J14.GZ43_509386 | AB008430 | Homo sapiens mRNA for CDEP, complete cds | 4.70E-05 |
| 476 | 3574.J23.GZ43_509530 | AP000384 | Arabidopsis thaliana genomic DNA, chromosome 3, P1 clone:MCE21 | 7.10E-07 |
| 477 | 3574.K12.GZ43_509355 | AB031814 | Mus musculus oatp2 mRNA for organic anion transporting polypeptide 2, complete cds | 1.50E-05 |
| 478 | 3574.K20.GZ43_509483 | AF126719 | Plasmodium falciparum cAMP-dependent protein kinase (pka) gene, complete cds | 3.00E-06 |
| 479 | 3574.L07.GZ43_509276 | U53400 | Rattus norvegicus chromosome 10 microsatellite sequence D10Mco21 | 8.94E-02 |
| 480 | 3574.M03.GZ43_509213 | AB000404 | Rice grassy stunt virus genomic RNA6 for 20.6K major nonstructural protein and 36.4K protein, complete cds | 5.60E-07 |
| 481 | 3574.M23.GZ43_509533 | U18056 | Lycopersicon esculentum 1-amino-cyclopropane-1-carboxylate synthase (LE-ACS1A) gene, complete cds | 3.40E-07 |
| 482 | 3574.N04.GZ43_509230 | L48479 | Homo sapiens (subclone 6_h1 from P1 H21) DNA sequence | 3.30E-09 |
| 483 | 3574.N10.GZ43_509326 | M58150 | Bovine lactoperoxidase (LPO) mRNA, complete cds | 3.60E-05 |
| 484 | 3574.N12.GZ43_509358 | AF182950 | Homo sapiens HEX (HEX) gene, partial cds and 5' flanking sequence | 9.00E-06 |
| 485 | 3574.N20.GZ43_509486 | AE006904 | Sulfolobus solfataricus section 263 of 272 of the complete genome | 3.00E-06 |
| 486 | 3574.P07.GZ43_509280 | U60232 | Homo sapiens cysteine dioxygenase (CDO-1) gene, 5' flanking region and exons 1 and 2 | 6.30E-08 |
| 487 | 3574.P17.GZ43_509440 | AC002218 | Homo sapiens (subclone 2_c1 from P1 H43) DNA sequence, complete sequence | 5.30E-08 |
| 488 | 3577.A06.GZ43_509633 | U28328 | Bos taurus dinucleotide repeat RM154, tandem repeat region | 3.40E-27 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 489 | 3577.A18.GZ43_509825 | X58774 | Herpesvirus saimiri sRNA1, sRNA2, sRNA3 and sRNA4 genes for small viral RNAs | 1.00E-06 |
| 490 | 3577.B12.GZ43_509730 | BC008400 | Homo sapiens, postmeiotic segregation increased (S. cerevisiae) 2, clone IMAGE:4273792, mRNA | 2.50E-05 |
| 491 | 3577.B15.GZ43_509778 | M61127 | Drosophila melanogaster GTP-binding protein (arf-like) gene, complete cds | 1.10E-05 |
| 492 | 3577.B19.GZ43_509842 | AF135526 | Homo sapiens clone MINT26 colon cancer differentially methylated CpG island genomic sequence | 1.00E-06 |
| 493 | 3577.E19.GZ43_509845 | AF063864 | Schizosaccharomyces pombe essential nuclear protein Mcm3p (mcm3+) gene, complete cds | 1.00E-06 |
| 494 | 3577.F02.GZ43_509574 | U37434 | Danio rerio L-isoaspartate (D-aspartate) O-methyltransferase (PCMT) mRNA, complete cds | 5.10E-08 |
| 495 | 3577.G07.GZ43_509655 | AF001893 | Human MEN1 region clone epsilon/beta mRNA, 3' fragment | 3.00E-06 |
| 496 | 3577.G13.GZ43_509751 | M83821 | Xenopus laevis mucin B.1 consensus repeat mRNA | 2.10E-07 |
| 497 | 3577.H06.GZ43_509640 | AK007565 | Mus musculus 10 day old male pancreas cDNA, RIKEN full-length enriched library, clone:1810020K22, full insert sequence | 8.00E-07 |
| 498 | 3577.H08.GZ43_509672 | L81912 | Homo sapiens (subclone 2_g5 from PAC H74) DNA sequence, complete sequence | 2.40E-07 |
| 499 | 3577.H18.GZ43_509832 | AL157461 | Homo sapiens mRNA; cDNA DKFZp434K152 (from clone DKFZp434K152) | 4.00E-06 |
| 500 | 3577.I01.GZ43_509561 | U35006 | Carcharhinus plumbeus Ig lambda light chain gene, complete cds | 2.00E-06 |
| 501 | 3577.I17.GZ43_509817 | AF157252 | Gongronella butleri translation elongation factor 1-alpha (EF-1alpha) gene, partial cds | 1.00E-06 |
| 502 | 3577.J04.GZ43_509610 | AF338249 | Sus scrofa thyroid-stimulating hormone receptor mRNA, complete cds | 2.00E-06 |
| 503 | 3577.K06.GZ43_509643 | AB000264 | Bacillus firmus DNA for beta-amylase, partial cds | 5.00E-07 |
| 504 | 3577.K14.GZ43_509771 | X15441 | Aspergillus nidulans mitochondrial ndhC and oxiB genes for NADH dehydrogenase subunit 3 and cytochrome oxidase subunit II | 1.00E-06 |
| 505 | 3577.K23.GZ43_509915 | X52952 | Rat mRNA for c-mos | 3.00E-06 |
| 506 | 3577.L10.GZ43_509708 | X60578 | Hepatitis C genomic RNA for putative envelope protein (RE56 isolate) | 3.70E-07 |
| 507 | 3577.N10.GZ43_509710 | Z75121 | S.cerevisiae chromosome XV reading frame ORF YOR213c | 4.50E-09 |
| 508 | 3577.N14.GZ43_509774 | M90058 | Human serglycin gene, exons 1,2, and 3 | 5.00E-06 |
| 509 | 3577.O17.GZ43_509823 | L19141 | Lupinus albus L-asparaginase gene, complete cds | 9.10E-08 |

**Table 8**

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 510 | 3577.O22.GZ43_509903 | AL031008 | Human DNA sequence from clone 360A4 on chromosome 16. Contains ESTs, complete sequence [Homo sapiens] | 5.60E-08 |
| 511 | 3577.P02.GZ43_509584 | AK006176 | Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:1700020M10, full insert sequence | 4.60E-08 |
| 512 | 3577.P07.GZ43_509664 | U05822 | Human proto-oncogene BCL3 gene, exon 2 | 2.40E-14 |
| 513 | 3577.P23.GZ43_509920 | AJ010341 | Homo sapiens PISSLRE gene, exons 1, 2, and 3 and joined CDS | 1.00E-11 |
| 514 | 3580.A04.GZ43_509985 | AJ010213 | Mus musculus beta-dystrobrevin gene, exon 10 | 8.20E-07 |
| 515 | 3580.A09.GZ43_510065 | AB037862 | Homo sapiens mRNA for KIAA1441 protein, partial cds | 6.30E-15 |
| 516 | 3580.A13.GZ43_510129 | U17832 | Symploce pallens mitochondrion 16S ribosomal RNA, partial sequence | 7.80E-07 |
| 517 | 3580.A14.GZ43_510145 | X89414 | A.thaliana DNA for pyrroline-5-carboxylase synthetase gene | 6.00E-06 |
| 518 | 3580.B01.GZ43_509938 | U67487 | Methanococcus jannaschii section 29 of 150 of the complete genome | 9.00E-05 |
| 519 | 3580.C01.GZ43_509939 | X14898 | Hamster p7 preinsertion DNA | 2.00E-06 |
| 520 | 3580.C03.GZ43_509971 | X76302 | H.sapiens RY-1 mRNA for putative nucleic acid binding protein | 3.70E-07 |
| 521 | 3580.C05.GZ43_510003 | Z22923 | M.musculus alpha2 (IX) collagen gene, complete CDS | 1.60E-05 |
| 522 | 3580.D07.GZ43_510036 | AB062941 | Macaca fascicularis brain cDNA clone:QflA-14927, full insert sequence | 9.80E-22 |
| 523 | 3580.D22.GZ43_510276 | M84136 | Flaveria chloraefolia flavonol 4'-sulfotransferase mRNA, complete cds | 4.00E-06 |
| 524 | 3580.E02.GZ43_509957 | AE001002 | Archaeoglobus fulgidus section 105 of 172 of the complete genome | 3.90E-05 |
| 525 | 3580.E08.GZ43_510053 | U48431 | Drosophila pseudoobscura alpha-amylase (Amy3) pseudogene, complete cds | 3.00E-06 |
| 526 | 3580.E10.GZ43_510085 | Z64717 | H.sapiens CpG island DNA genomic Msel fragment, clone 161e9, forward read cpg161e9.ft1a | 9.60E-19 |
| 527 | 3580.E19.GZ43_510229 | M64984 | Candida tropicalis open reading frame DNA sequence | 2.00E-06 |
| 528 | 3580.E21.GZ43_510261 | M84136 | Flaveria chloraefolia flavonol 4'-sulfotransferase mRNA, complete cds | 5.00E-06 |
| 529 | 3580.E23.GZ43_510293 | AB033570 | Eptatretus burgeri hgPTPR5a mRNA, partial cds | 2.00E-06 |
| 530 | 3580.G03.GZ43_509975 | Y14277 | Drosophila melanogaster mRNA for nuclear protein SA | 1.10E-05 |
| 531 | 3580.G13.GZ43_510135 | AK018491 | Mus musculus adult male colon cDNA, RIKEN full-length enriched library, clone:9030408N04, full insert sequence | 4.40E-08 |
| 532 | 3580.G14.GZ43_510151 | AF142660 | Lama glama microsatellite LCA90 sequence | 2.60E-07 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 533 | 3580.G18.GZ43_510215 | D86226 | Spinacia oleracea DNA for nitrate reductase, complete cds | 2.60E-05 |
| 534 | 3580.G19.GZ43_510231 | U60502 | Glycine max actin (Soy119) gene, partial cds | 7.00E-06 |
| 535 | 3580.G20.GZ43_510247 | D38524 | Human mRNA for 5'-nucleotidase | 4.80E-11 |
| 536 | 3580.G24.GZ43_510311 | AF084480 | Mus musculus Williams-Beuren syndrome deletion transcript 9 homolog (Wbscr9) mRNA, complete cds | 5.00E-06 |
| 537 | 3580.H12.GZ43_510120 | X78423 | D.carota (Queen Anne's Lace) Inv*Dc3 gene, 4444bp | 4.00E-06 |
| 538 | 3580.H16.GZ43_510184 | Y13786 | Homo sapiens mRNA for meltrin-beta/ADAM 19 homologue | 4.50E-10 |
| 539 | 3580.H22.GZ43_510280 | X62578 | C.caldarium plastid genes ompR', psbD, psbC, rps16 and groEL | 2.50E-05 |
| 540 | 3580.I06.GZ43_510025 | X51344 | Spiroplasma virus (SpV1-R8A2 B) complete genome | 4.70E-07 |
| 541 | 3580.I08.GZ43_510057 | X02761 | Human mRNA for fibronectin (FN precursor) | 1.02E-04 |
| 542 | 3580.I18.GZ43_510217 | BC007856 | Homo sapiens, clone MGC:14337 IMAGE:4298428, mRNA, complete cds | 2.60E-10 |
| 543 | 3580.J10.GZ43_510090 | AF068206 | Rangifer tarandus microsatellite NVHRT16 sequence | 4.40E-11 |
| 544 | 3580.J12.GZ43_510122 | AE008323 | Agrobacterium tumefaciens strain C58 linear chromosome, section 127 of 187 of the complete sequence | 9.30E-05 |
| 545 | 3580.J18.GZ43_510218 | AF222689 | Homo sapiens protein arginine N-methyltransferase 1 (HRMT1L2) gene, complete cds, alternatively spliced | 1.50E-05 |
| 546 | 3580.J20.GZ43_510250 | M31651 | Homo sapiens sex hormone-binding globulin (SHBG) gene, complete cds | 3.80E-07 |
| 547 | 3580.J21.GZ43_510266 | AB054062 | Pagrus major lpl mRNA for lipoprotein lipase, complete cds | 3.00E-06 |
| 548 | 3580.K03.GZ43_509979 | AE007607 | Clostridium acetobutylicum ATCC824 section 95 of 356 of the complete genome | 5.00E-05 |
| 549 | 3580.K05.GZ43_510011 | Z15027 | H.sapiens HLA class III DNA | 3.70E-08 |
| 550 | 3580.K21.GZ43_510267 | AF135826 | Mus musculus neuronal nitric oxide synthase (NOS-I) gene, exon 1c and 5'-flanking sequence | 2.20E-09 |
| 551 | 3580.L09.GZ43_510076 | AL049333 | Homo sapiens mRNA; cDNA DKFZp564M116 (from clone DKFZp564M116) | 3.40E-13 |
| 552 | 3580.L10.GZ43_510092 | AF278587 | Borrelia burgdorferi strain BC-1 outer surface protein C (ospC) gene, partial cds | 2.00E-06 |
| 553 | 3580.L12.GZ43_510124 | D14664 | Human mRNA for KIAA0022 gene, complete cds | 1.10E-05 |
| 554 | 3580.L13.GZ43_510140 | K02269 | Human ERV3 (endogenous retrovirus 3) gag gene | 3.30E-07 |
| 555 | 3580.L17.GZ43_510204 | U60232 | Homo sapiens cysteine dioxygenase (CDO-1) gene, 5' flanking region and exons 1 and 2 | 2.00E-07 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 556 | 3580.M01.GZ43_509949 | U53400 | Rattus norvegicus chromosome 10 microsatellite sequence D10Mco21 | 4.54E-01 |
| 557 | 3580.M16.GZ43_510189 | AE006406 | Lactococcus lactis subsp. lactis IL1403 section 168 of 218 of the complete genome | 3.00E-06 |
| 558 | 3580.M17.GZ43_510205 | AF348584 | Arabidopsis thaliana unknown protein (T8K14.7) mRNA, complete cds | 6.70E-07 |
| 559 | 3580.M18.GZ43_510221 | X69908 | H.sapiens gene for mitochondrial ATP synthase c subunit (P2 form) | 1.00E-05 |
| 560 | 3580.M23.GZ43_510301 | M17326 | Mouse endogenous murine leukemia virus polytropic provirus DNA, complete cds | 9.00E-06 |
| 561 | 3580.N10.GZ43_510094 | AF103970 | Lasioglossum rohweri cytochrome oxidase I (COI) gene, mitochondrial gene encoding mitochondrial protein, partial cds | 1.00E-06 |
| 562 | 3580.N11.GZ43_510110 | Z80362 | H.sapiens HLA-DRB pseudogene, exon 1; | 6.10E-11 |
| 563 | 3580.N14.GZ43_510158 | AB014462 | Xenopus laevis XNLRR-1 mRNA, complete cds | 1.60E-05 |
| 564 | 3580.N15.GZ43_510174 | AF164381 | Anomochloa marantoidea maturase (matK) gene, complete cds; chloroplast gene for chloroplast product | 1.00E-06 |
| 565 | 3580.N23.GZ43_510302 | AB047880 | Macaca fascicularis brain cDNA, clone:QnpA-14303 | 2.00E-06 |
| 566 | 3580.O02.GZ43_509967 | X55948 | H. aspersa cytoplasmic intermediate filament gene exons 2 to 6 | 4.00E-06 |
| 567 | 3580.O06.GZ43_510031 | L34649 | Homo sapiens platelet/endothelial cell adhesion molecule-1 (PECAM-1) gene, exon 14 | 4.00E-06 |
| 568 | 3580.O07.GZ43_510047 | Z30183 | H.sapiens mig-5 gene | 3.00E-05 |
| 569 | 3580.O08.GZ43_510063 | AF101385 | Homo sapiens ribosomal protein L11 gene, complete cds | 1.80E-08 |
| 570 | 3580.P04.GZ43_510000 | AC016707 | Homo sapiens BAC clone RP11-221K4 from Y, complete sequence | 1.80E-08 |
| 571 | 3580.P05.GZ43_510016 | AF055482 | Thermotoga neapolitana galactose utilization operon, complete sequence | 8.00E-07 |
| 572 | 3580.P14.GZ43_510160 | AF009133 | Rattus norvegicus CD94 (Cd94) mRNA, complete cds | 7.50E-08 |
| 573 | 3580.P19.GZ43_510240 | Y15176 | Human papillomavirus type 80 E6, E7, E1, E2, E4, L2, and L1 genes | 7.00E-06 |
| 574 | 3583.B06.GZ43_510402 | X51398 | Chlamydomonas moewusii chloroplast DNA for ORF 563 and transfer RNA-Thr | 3.00E-06 |
| 575 | 3583.B07.GZ43_510418 | U39382 | Hexachaeta amabilis 16S ribosomal RNA gene, mitochondrial gene encoding mitochondrial RNA, partial sequence | 5.50E-08 |
| 576 | 3583.B10.GZ43_510466 | S45332 | erythropoietin receptor [human, placental, Genomic, 8647 nt] | 3.90E-10 |
| 577 | 3583.B11.GZ43_510482 | AC006623 | Caenorhabditis elegans clone C52E2, complete sequence | 4.00E-06 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 578 | 3583.D15.GZ43_510548 | AF242297 | Homo sapiens phosducin-like protein gene, promoter and exon 1 | 3.80E-08 |
| 579 | 3583.D22.GZ43_510660 | Z23548 | H. sapiens (D10S540) DNA segment containing (CA) repeat; clone AFM205xe11; single read | 3.20E-07 |
| 580 | 3583.E11.GZ43_510485 | X69737 | E.esula chloroplast rbcL gene for ribulose-1,5-biphosphate-carboxylase and promoter region | 1.30E-08 |
| 581 | 3583.E13.GZ43_510517 | AB007856 | Homo sapiens KIAA0396 mRNA, partial cds | 2.20E-05 |
| 582 | 3583.E15.GZ43_510549 | X74131 | H.nelsoni small subunit ribosomal RNA | 7.00E-06 |
| 583 | 3583.E17.GZ43_510581 | AE006633 | Streptococcus pyogenes M1 GAS strain SF370, section 162 of 167 of the complete genome | 2.40E-07 |
| 584 | 3583.F24.GZ43_510694 | J02846 | Human tissue factor gene, complete cds | 7.40E-07 |
| 585 | 3583.G09.GZ43_510455 | X88789 | P.sativum mRNA for starch synthase (2035 bp) | 2.10E-05 |
| 586 | 3583.G16.GZ43_510567 | AK000735 | Homo sapiens cDNA FLJ20728 fis, clone HEP11763 | 4.70E-07 |
| 587 | 3583.G17.GZ43_510583 | AK026822 | Homo sapiens cDNA: FLJ23169 fis, clone LNG09957 | 2.60E-05 |
| 588 | 3583.G21.GZ43_510647 | U13044 | Human nuclear respiratory factor-2 subunit alpha mRNA, complete cds | 2.00E-06 |
| 589 | 3583.H03.GZ43_510360 | M26222 | African green monkey origin of replication (ORS9) region | 1.00E-13 |
| 590 | 3583.H12.GZ43_510504 | X01669 | Human c-k-ras oncogene exon 2 from lung carcinoma pr310 | 3.20E-08 |
| 591 | 3583.H13.GZ43_510520 | AK022380 | Homo sapiens cDNA FLJ12318 fis, clone MAMMA1002068 | 2.00E-06 |
| 592 | 3583.H15.GZ43_510552 | L77119 | Methanococcus jannaschii small extra-chromosomal element, complete sequence.~ | 1.60E-05 |
| 593 | 3583.J02.GZ43_510346 | AJ007302 | Sus scrofa triadin gene | 1.00E-06 |
| 594 | 3583.K08.GZ43_510443 | D63902 | Mouse mRNA for estrogen-responsive finger protein, complete cds | 2.50E-11 |
| 595 | 3583.K10.GZ43_510475 | U11816 | Lactobacillus strain 30A ornithine decarboxylase (odci) gene, complete cds | 1.00E-05 |
| 596 | 3583.K11.GZ43_510491 | X73416 | W.suaveolens mitochondrial orf1 | 6.00E-06 |
| 597 | 3583.K14.GZ43_510539 | U04367 | Bacillus thuringiensis dakota HD511 CryIII delta-endotoxin gene, partial cds | 1.20E-05 |
| 598 | 3583.K17.GZ43_510587 | AE004129 | Vibrio cholerae chromosome I, section 37 of 251 of the complete chromosome | 8.00E-06 |
| 599 | 3583.K23.GZ43_510683 | AE001410 | Plasmodium falciparum chromosome 2, section 47 of 73 of the complete sequence | 4.00E-06 |
| 600 | 3583.L05.GZ43_510396 | X55299 | C.stercorarium celZ gene for endo-beta-1,4-glucanase (Avicelase I) | 1.00E-05 |
| 601 | 3583.L08.GZ43_510444 | AF106953 | Homo sapiens SOS1 (SOS1) gene, partial cds | 7.50E-09 |
| 602 | 3583.L09.GZ43_510460 | L34842 | Soybean chloroplast phytochrome A (phyA) gene, complete cds | 2.40E-05 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 603 | 3583.L17.GZ43_510588 | X65223 | T.rubrum mitochondrion genes for cytochrome oxidase I, cytochrome oxidase II, ATPase 9, NADH dehydrogenase subunit 4L, NADH dehydrogenase subunit 5, tRNA-Gln, tRNA-Met and tRNA-Arg | 5.00E-06 |
| 604 | 3583.L21.GZ43_510652 | AF106661 | Rattus norvegicus glutathione S-transferase Yb4 (GstYb4) gene, complete cds | 5.00E-06 |
| 605 | 3583.M08.GZ43_510445 | BC005276 | Homo sapiens, Similar to GRO2 oncogene, clone IMAGE:4071652, mRNA | 3.70E-07 |
| 606 | 3583.M10.GZ43_510477 | Y00477 | Human bone marrow serine protease gene (medullasin) (leukocyte neutrophil elastase gene) | 4.70E-09 |
| 607 | 3583.M13.GZ43_510525 | X73030 | S.cerevisiae YGP1 gene | 7.00E-06 |
| 608 | 3583.N09.GZ43_510462 | AK018377 | Mus musculus 16 days embryo lung cDNA, RIKEN full-length enriched library, clone:8430403M08, full insert sequence | 4.60E-07 |
| 609 | 3583.O03.GZ43_510367 | X72698 | P.pygmaeus ZFY gene for Y-linked Zinc finger protein, final intron | 3.00E-06 |
| 610 | 3583.O11.GZ43_510495 | U40161 | Arabidopsis thaliana type 2A protein serine/threonine phosphatase 55 kDa B regulatory subunit mRNA, complete cds | 2.00E-06 |
| 611 | 3583.O17.GZ43_510591 | U67567 | Methanococcus jannaschii section 109 of 150 of the complete genome | 2.00E-06 |
| 612 | 3583.P09.GZ43_510464 | AK021312 | Mus musculus 13 days embryo stomach cDNA, RIKEN full-length enriched library, clone:D530039A21, full insert sequence | 3.60E-08 |
| 613 | 3583.P19.GZ43_510624 | U12920 | Caenorhabditis elegans sex determination (tra-3) gene, exons 2-6 | 1.60E-05 |
| 614 | 3583.P22.GZ43_510672 | AJ133800 | Homo sapiens CPNE7 gene (partial), exon 2 | 7.60E-07 |
| 615 | 3590.A12.GZ43_512274 | AF185661 | Glomus intraradices strain FL208 18S ribosomal RNA, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA and internal transcribed spacer 2, complete sequence; 26S ribosomal RNA, partial sequence | 2.00E-06 |
| 616 | 3590.B01.GZ43_512099 | M96068 | Madagascar periwinkle hydroxymethylglutaryl-CoA reductase (HMGR) mRNA, complete cds | 7.40E-09 |
| 617 | 3590.B16.GZ43_512339 | V01527 | Mouse gene coding for major histocompatibility antigen. This is a class II antigen, I-A-beta | 2.40E-12 |
| 618 | 3590.B21.GZ43_512419 | AB028983 | Homo sapiens mRNA for KIAA1060 protein, partial cds | 1.70E-05 |
| 619 | 3590.C20.GZ43_512404 | D86566 | Human DNA for NOTCH4, partial cds | 3.20E-07 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 620 | 3590.D03.GZ43_512133 | D10371 | Phocine distemper virus (PDV) genomic RNA for N, P, V, C, M, F, H and L protein | 2.90E-05 |
| 621 | 3590.D19.GZ43_512389 | M96163 | Mus musculus (clone 2) serum inducible kinase (SNK) mRNA, mRNA sequence | 7.80E-10 |
| 622 | 3590.D23.GZ43_512453 | AF086485 | Homo sapiens full length insert cDNA clone ZD93E02 | 7.70E-09 |
| 623 | 3590.E08.GZ43_512214 | AF055278 | Homo sapiens DNA repair protein XRCC4 (XRCC4) gene, exon 1 | 5.90E-12 |
| 624 | 3590.E10.GZ43_512246 | AE001477 | Helicobacter pylori, strain J99 section 38 of 132 of the complete genome | 2.00E-06 |
| 625 | 3590.F01.GZ43_512103 | AF080395 | Entamoeba histolytica actin binding protein (abp2) mRNA, partial cds | 2.00E-06 |
| 626 | 3590.F16.GZ43_512343 | X79388 | B.subtilis (168) prkA gene | 1.20E-05 |
| 627 | 3590.G01.GZ43_512104 | U32690 | Haemophilus influenzae Rd section 5 of 163 of the complete genome | 2.80E-05 |
| 628 | 3590.G02.GZ43_512120 | U68040 | Cochliobolus heterostrophus polyketide synthase (PKS1) gene, complete cds | 1.25E-04 |
| 629 | 3590.H04.GZ43_512153 | X66013 | T.aestivum gene for cathepsin B (Al16) | 2.50E-07 |
| 630 | 3590.H06.GZ43_512185 | X66177 | M.musculus mRNA for Hox 2.7 protein | 8.00E-06 |
| 631 | 3590.H09.GZ43_512233 | AF012899 | Sambucus nigra ribosome inactivating protein precursor mRNA, complete cds | 3.40E-11 |
| 632 | 3590.H12.GZ43_512281 | Y15724 | Homo sapiens SERCA3 gene, exons 1-7 (and joined CDS) | 2.00E-06 |
| 633 | 3590.H16.GZ43_512345 | AF064079 | Plasmodium gallinaceum endochitinase precursor, mRNA, complete cds | 6.70E-09 |
| 634 | 3590.I16.GZ43_512346 | L06280 | Drosophila melanogaster adenine phosphoribosyltransferase (APRT) gene, complete cds | 4.40E-07 |
| 635 | 3590.J01.GZ43_512107 | X69573 | T.reesei xyn1 gene, complete CDS | 1.70E-07 |
| 636 | 3590.J02.GZ43_512123 | AF092047 | Homo sapiens homeobox protein Six3 (SIX3) gene, complete cds | 4.00E-06 |
| 637 | 3590.J18.GZ43_512379 | AB027966 | Schizosaccharomyces pombe gene for Hypothetical protein, partial cds, clone:TB89 | 2.60E-08 |
| 638 | 3590.J21.GZ43_512427 | AK014727 | Mus musculus 0 day neonate head cDNA, RIKEN full-length enriched library, clone:4833419G08, full insert sequence | 7.90E-08 |
| 639 | 3590.J22.GZ43_512443 | AK020136 | Mus musculus 12 days embryo male wolffian duct includes surrounding region cDNA, RIKEN full-length enriched library, clone:6720460K10, full insert sequence | 5.90E-08 |
| 640 | 3590.K06.GZ43_512188 | AF171890 | Trimeresurus trigonocephalus cytochrome b (cytb) gene, partial cds; mitochondrial gene for mitochondrial product | 3.00E-06 |
| 641 | 3590.K10.GZ43_512252 | U16775 | Human immunodeficiency virus type 1 isolate VE6 reverse transcriptase (pol) gene, partial cds | 6.00E-06 |

**Table 8**

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 642 | 3590.K19.GZ43_512396 | U40454 | Candida albicans topoisomerase type I (CATOP1) gene, complete cds | 3.00E-06 |
| 643 | 3590.L08.GZ43_512221 | U52198 | Vibrio anguillarum flagellin E (flaE), flagellin D (flaD), and flagellin B (flaB) genes, complete cds, and (flaG) gene, partial cds | 2.00E-05 |
| 644 | 3590.L10.GZ43_512253 | U01155 | Xenopus laevis angiotensin II receptor mRNA, complete cds | 4.00E-06 |
| 645 | 3590.M03.GZ43_512142 | AF252499 | Bos taurus clone MNB-88 microsatellite sequence | 4.60E-08 |
| 646 | 3590.M04.GZ43_512158 | AE007607 | Clostridium acetobutylicum ATCC824 section 95 of 356 of the complete genome | 4.50E-05 |
| 647 | 3590.M09.GZ43_512238 | L04758 | Oryctolagus cuniculus cytochrome P-450 (CYP4A4) gene, 5' end | 1.00E-06 |
| 648 | 3590.N04.GZ43_512159 | Z82038 | C.thermosaccharolyticum etfB, etfA, hbd, thlA and actA genes | 2.00E-06 |
| 649 | 3590.N19.GZ43_512399 | U15603 | Saccharomyces cerevisiae Csd3p (CSD3) gene, complete cds | 4.00E-06 |
| 650 | 3590.N21.GZ43_512431 | L19535 | Drosophila subobscura sry alpha gene, complete cds | 6.00E-06 |
| 651 | 3590.O08.GZ43_512224 | L36588 | Homo sapiens intron-encoded U22 small nucleolar RNA (UHG) gene | 4.30E-07 |
| 652 | 3596.C02.GZ43_512500 | L14849 | Drosophila melanogaster cytoplasmic protein tyrosine phosphatase (PTP61F) mRNA, complete cds | 8.90E-09 |
| 653 | 3596.C20.GZ43_512788 | M60286 | Herpesvirus saimiri immediate early region protein genes, complete cds | 1.30E-07 |
| 654 | 3596.C22.GZ43_512820 | X15121 | Soybean Gy1 gene for glycinin subunit G1 | 1.00E-06 |
| 655 | 3596.D01.GZ43_512485 | Z78414 | Caenorhabditis elegans cosmid W09D12, complete sequence | 4.00E-06 |
| 656 | 3596.D07.GZ43_512581 | M88242 | Mouse glucocortoid-regulated inflammatory prostaglandin G/H synthase (griPGHS) mRNA, complete cds | 1.70E-05 |
| 657 | 3596.D09.GZ43_512613 | X99710 | L.lactis ORF, genes homologous to vsf-1 and pepF2 and gene encoding protein homologous to methyltransferase | 5.00E-06 |
| 658 | 3596.D17.GZ43_512741 | AF200361 | Rattus norvegicus cytochrome P450 4F1 (Cyp4F1) gene, complete cds | 1.40E-05 |
| 659 | 3596.E08.GZ43_512598 | AF111848 | Homo sapiens PRO0529 mRNA, complete cds | 5.00E-06 |
| 660 | 3596.E22.GZ43_512822 | X58178 | S.pyogenes for emm41 gene | 5.00E-06 |
| 661 | 3596.F10.GZ43_512631 | AL390161 | Homo sapiens mRNA; cDNA DKFZp761P0615 (from clone DKFZp761P0615) | 2.00E-06 |
| 662 | 3596.G13.GZ43_512680 | AJ000044 | Tenebrio molitor LPCP29 gene | 2.00E-06 |
| 663 | 3596.H04.GZ43_512537 | U65018 | Dictyostelium discoideum mannosyltransferase gene, complete cds | 3.60E-07 |
| 664 | 3596.H10.GZ43_512633 | AF104390 | Penaeus monodon hyperglycemic hormone homolog PmSGP-V precursor, mRNA, complete cds | 2.00E-06 |

**Table 8**

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 665 | 3596.H17.GZ43_512745 | D28915 | Human gene for hepatitis C-associated microtubular aggregate protein p44, exon 9 and complete cds | 1.00E-06 |
| 666 | 3596.H22.GZ43_512825 | AF198250 | Dictyostelium discoideum lim2 protein (limB) mRNA, complete cds | 7.30E-07 |
| 667 | 3596.I06.GZ43_512570 | U32444 | Solanum lycopersicum phytochrome F (PHYF) gene, partial cds | 1.10E-05 |
| 668 | 3596.I16.GZ43_512730 | U32444 | Solanum lycopersicum phytochrome F (PHYF) gene, partial cds | 8.00E-06 |
| 669 | 3596.J04.GZ43_512539 | D28596 | Chicken gene for c-maf proto-oncogene product c-Maf, short form complete cds and long form 1st exon | 9.30E-10 |
| 670 | 3596.J13.GZ43_512683 | AB007856 | Homo sapiens KIAA0396 mRNA, partial cds | 2.40E-05 |
| 671 | 3596.K14.GZ43_512700 | AC024752 | Caenorhabditis elegans cosmid Y1B5A, complete sequence | 3.00E-06 |
| 672 | 3596.K15.GZ43_512716 | Y00469 | Yeast mRNA for profilin | 2.00E-06 |
| 673 | 3596.L01.GZ43_512493 | X79703 | O.aries gene for beta-casein | 4.00E-06 |
| 674 | 3596.L08.GZ43_512605 | AJ007313 | Streptomyces coelicolor sigT, trxB and trxA genes, and ORF1 and ORF2 | 9.80E-07 |
| 675 | 3596.L13.GZ43_512685 | AK018239 | Mus musculus adult male medulla oblongata cDNA, RIKEN full-length enriched library, clone:6330563C09, full insert sequence | 1.00E-06 |
| 676 | 3596.N02.GZ43_512511 | AE001387 | Plasmodium falciparum chromosome 2, section 24 of 73 of the complete sequence | 1.00E-06 |
| 677 | 3596.N12.GZ43_512671 | Z12841 | O.cuniculus mRNA for phospholipase | 4.00E-06 |
| 678 | 3596.N15.GZ43_512719 | U14186 | Bos taurus general vesicular transport factor p115 mRNA, complete cds | 1.70E-05 |
| 679 | 3596.N16.GZ43_512735 | U41106 | Caenorhabditis elegans cosmid W06A11 | 1.10E-05 |
| 680 | 3596.N21.GZ43_512815 | AF097717 | Homo sapiens 3'-phosphoadenosine 5'-phosphosulfate synthetase (PAPSS), exon 8 | 1.40E-07 |
| 681 | 3596.O10.GZ43_512640 | AE001649 | Chlamydia pneumoniae section 65 of 103 of the complete genome | 1.10E-05 |
| 682 | 3596.O12.GZ43_512672 | AC006623 | Caenorhabditis elegans clone C52E2, complete sequence | 4.00E-06 |
| 683 | 3596.P03.GZ43_512529 | X82317 | C.thummi CpY gene | 1.49E-03 |
| 684 | 3596.P04.GZ43_512545 | AF111855 | Agrobacterium tumefaciens RNA polymerase alpha subunit (rpoA) gene, complete cds | 2.00E-06 |
| 685 | 3596.P07.GZ43_512593 | L40817 | Homo sapiens muscle-specific DNase I-like (DNL1L) gene, exons 1-9, complete cds | 3.00E-06 |
| 686 | 3596.P08.GZ43_512609 | M14505 | Human (clone PSK-J3) cyclin-dependent protein kinase mRNA, complete cds., | 5.00E-06 |
| 687 | 3596.P10.GZ43_512641 | M73770 | P.falciparum RNA polymerase III largest subunit gene, complete cds | 2.90E-05 |

**Table 8**

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 688 | 3596.P21.GZ43_512817 | S82725 | NPM/ALK=fusion gene {translocation breakpoint} [human, lymphoma cells SU-DHL-1, Genomic, 1679 nt] | 1.00E-07 |
| 689 | 3599.A04.GZ43_512914 | X83212 | H.sapiens tryptophan hydroxylase gene, promoter region | 5.50E-07 |
| 690 | 3599.A23.GZ43_513218 | U05259 | Human MB-1 gene, complete cds | 2.10E-05 |
| 691 | 3599.B15.GZ43_513091 | AF277068 | HIV-1 clone QH0791 from Trinidad and Tobago, envelope protein (env) gene, complete cds | 6.10E-07 |
| 692 | 3599.B16.GZ43_513107 | M60517 | Chicken vitronectin receptor alpha subunit mRNA, complete cds | 4.00E-06 |
| 693 | 3599.C03.GZ43_512900 | AB021267 | Arabidopsis thaliana copia-like retrotransposon AtRE2-2 gene for polyprotein, complete cds | 2.00E-06 |
| 694 | 3599.C17.GZ43_513124 | U28055 | Homo sapiens hepatocyte growth factor-like protein homolog mRNA, partial cds | 3.00E-06 |
| 695 | 3599.D03.GZ43_512901 | L43550 | Buchnera aphidicola anthranilate synthase small subunit (trpG) gene, anthranilate synthase large subunit (trpE) gene, complete cds | 3.00E-06 |
| 696 | 3599.D05.GZ43_512933 | AL023779 | S.pombe chromosome II cosmid c244 | 2.00E-06 |
| 697 | 3599.D07.GZ43_512965 | AL391223 | Human chromosome 14 DNA sequence Partial sequence from BAC R-325N7_PCR1 of library RPCI-11 from chromosome 14 of Homo sapiens (Human), complete sequence | 5.00E-06 |
| 698 | 3599.D10.GZ43_513013 | AF064079 | Plasmodium gallinaceum endochitinase precursor, mRNA, complete cds | 1.70E-07 |
| 699 | 3599.E01.GZ43_512870 | U09184 | Buchnera aphidicola ferredoxin-NADP reductase (fprl) gene, partial cds; anthranilate synthase large subunit (trpE) and anthranilate synthase small subunit (trpG) genes, complete cds; heat shock protein (hslU) gene, partial cds; and unknown gene | 9.60E-07 |
| 700 | 3599.E05.GZ43_512934 | X60145 | Human J-alpha segment J-alpha FR9 mRNA for J-alpha region of T-cell receptor | 1.20E-05 |
| 701 | 3599.F17.GZ43_513127 | U27037 | Fistulina hepatica mitochondrial small subunit ribosomal RNA, mitochondrial gene, partial sequence | 2.00E-06 |
| 702 | 3599.F24.GZ43_513239 | Z78414 | Caenorhabditis elegans cosmid W09D12, complete sequence | 5.00E-06 |
| 703 | 3599.H05.GZ43_512937 | AF032891 | Camponotus consobrinus microsatellite-containing sequence Ccon12 | 2.10E-08 |
| 704 | 3599.H23.GZ43_513225 | AB024553 | Bacillus halodurans DNA, complete and partial cds, strain:C-125 | 4.70E-07 |
| 705 | 3599.J11.GZ43_513035 | AB025112 | Xenopus laevis XGC-2 mRNA for guanylyl cyclase-2, complete cds | 3.00E-06 |
| 706 | 3599.K02.GZ43_512892 | AJ224474 | Borrelia burgdorferi left chromosomal subtelomeric region (truA gene) | 3.00E-06 |

**Table 8**

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 707 | 3599.K04.GZ43_512924 | X99710 | L.lactis ORF, genes homologous to vsf-1 and pepF2 and gene encoding protein homologous to methyltransferase | 5.00E-06 |
| 708 | 3599.K23.GZ43_513228 | AF074247 | Homo sapiens neuronal delayed-rectifier voltage-gated potassium channel splice variant (KCNQ2) mRNA, complete cds | 8.00E-07 |
| 709 | 3599.L04.GZ43_512925 | X59773 | Pisum sativum mRNA for P protein, a part of glycine cleavage complex | 1.40E-05 |
| 710 | 3599.L15.GZ43_513101 | U34282 | Rattus norvegicus fast skeletal muscle sarcoplasmic reticulum Ca-ATPase (SERCA1) gene, 5'-flanking sequence | 2.00E-06 |
| 711 | 3599.M04.GZ43_512926 | AK018953 | Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:1700111D04, full insert sequence | 2.30E-11 |
| 712 | 3599.M22.GZ43_513214 | AB052179 | Macaca fascicularis brain cDNA, clone:QnpA-21934 | 4.70E-07 |
| 713 | 3599.M24.GZ43_513246 | AE003394 | Drosophila melanogaster genomic scaffold 142000013386028, complete sequence | 7.30E-07 |
| 714 | 3599.N09.GZ43_513007 | X16362 | Rat SPI-2 serine protease inhibitor gene | 1.19E-04 |
| 715 | 3599.N16.GZ43_513119 | X92421 | X.laevis mRNA for RNA helicase p54 | 3.00E-06 |
| 716 | 3599.N20.GZ43_513183 | M59447 | Drosophila melanogaster Sex-lethal (Sxl) mRNA, complete cds | 2.00E-06 |
| 717 | 3599.N24.GZ43_513247 | AC005485 | Homo sapiens PAC clone RP5-998M2 from 7q33-q35, complete sequence | 2.00E-07 |
| 718 | 3599.O06.GZ43_512960 | AJ131667 | Escherichia coli plasmid pSFO157 | 2.00E-06 |
| 719 | 3599.O17.GZ43_513136 | X96607 | M.musculus IgH 3' alpha enhancer DNA | 8.10E-05 |
| 720 | 3599.P05.GZ43_512945 | X77111 | N.tabacum chi-V gene | 1.50E-07 |
| 721 | 3602.A09.GZ43_513378 | AF015303 | Xenopus laevis small GTPase Ran binding protein 1 mRNA, complete cds | 1.10E-05 |
| 722 | 3602.B18.GZ43_513523 | L18892 | Tetrahymena thermophila histone (H2A.1) gene, complete cds | 5.70E-07 |
| 723 | 3602.B21.GZ43_513571 | BC005233 | Homo sapiens, clone MGC:12257 IMAGE:3950129, mRNA, complete cds | 1.60E-10 |
| 724 | 3602.B22.GZ43_513587 | X71765 | P. falciparum gene for Ca2+ - ATPase | 1.00E-06 |
| 725 | 3602.C24.GZ43_513620 | AL080106 | Homo sapiens mRNA; cDNA DKFZp566O053 (from clone DKFZp566O053) | 2.00E-06 |
| 726 | 3602.D06.GZ43_513333 | AF098970 | Phaseolus vulgaris NBS-LRR-like protein cD7 (CO-2) mRNA, partial cds | 1.70E-07 |
| 727 | 3602.D11.GZ43_513413 | M59770 | P.falciparum calmodulin gene, complete cds | 2.20E-07 |
| 728 | 3602.E04.GZ43_513302 | X53582 | Zea mays ZMPMS1 gene for 19 kDa zein protein | 1.30E-05 |
| 729 | 3602.E06.GZ43_513334 | L38718 | Providencia stuartii (clone pSK.aarP) transcriptional activator (aarP) gene, complete cds | 7.90E-07 |
| 730 | 3602.E13.GZ43_513446 | U58106 | Blomia tropicalis allergen mRNA, complete cds | 1.70E-07 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 731 | 3602.E21.GZ43_513574 | M15085 | T.brucei expressed copy of the ILTat 1.3 variable surface glycoprotein gene, 5' flank | 2.90E-07 |
| 732 | 3602.F12.GZ43_513431 | X64802 | H.sapiens F8 mRNA for Interleukin-1-like species | 3.40E-58 |
| 733 | 3602.G03.GZ43_513288 | AF036148 | Danio rerio NeuroD (nrd) mRNA, complete cds | 2.00E-06 |
| 734 | 3602.G17.GZ43_513512 | U41106 | Caenorhabditis elegans cosmid W06A11 | 1.30E-05 |
| 735 | 3602.I07.GZ43_513354 | AF000941 | Mus musculus DNAse I hypersensitive sites 2-6 of locus control region (LCR) for T-cell receptor alpha chain (TCRa) gene | 1.20E-05 |
| 736 | 3602.I11.GZ43_513418 | AL133620 | Homo sapiens mRNA; cDNA DKFZp434F0621 (from clone DKFZp434F0621) | 3.00E-06 |
| 737 | 3602.I15.GZ43_513482 | U23479 | Dictyostelium discoideum phosphatidylinositol 4-kinase (PIK4) mRNA, complete cds | 8.00E-07 |
| 738 | 3602.J13.GZ43_513451 | AK025319 | Homo sapiens cDNA: FLJ21666 fis, clone COL08915 | 3.30E-07 |
| 739 | 3602.K03.GZ43_513292 | X85811 | S.cerevisiae tRNA-Leu, and ORF's N2212, N2215, N2219, N2223, N2227, N2231 | 1.10E-05 |
| 740 | 3602.K06.GZ43_513340 | AF133052 | Walleye epidermal hyperplasia virus type 2 long terminal repeat, complete sequence; gag polyprotein (gag-pol) gene, complete cds; pol polyprotein (gag-pol) gene, partial cds; envelope polyprotein (env) and cyclin D homolog genes, complete cds; and unkn> | 4.00E-06 |
| 741 | 3602.L20.GZ43_513565 | M62717 | Human CSP-B gene flanking sequence | 1.10E-05 |
| 742 | 3602.N03.GZ43_513295 | Z81126 | Caenorhabditis elegans cosmid T22E6, complete sequence | 5.70E-05 |
| 743 | 3602.N06.GZ43_513343 | U62503 | Human OBR gene, intron sequence immediately adjacent to the 5' end of coding exon 17 | 1.00E-06 |
| 744 | 3605.A15.gz43_513858 | Z46507 | Bovine herpesvirus type 4 genomic DNA region (V.TEST) | 5.00E-06 |
| 745 | 3605.C16.gz43_513876 | AF282517 | Homo sapiens clone 10ptel_c6t7 sequence | 9.40E-08 |
| 746 | 3605.E19.gz43_513926 | Z22923 | M.musculus alpha2 (IX) collagen gene, complete CDS | 2.10E-05 |
| 747 | 3605.G13.gz43_513832 | AJ132752 | Gadus morhua mRNA for beta2-microglobulin, clone b3 | 1.30E-05 |
| 748 | 3605.H10.gz43_513785 | AF257480 | Rana temporaria microsatellite SB80 sequence | 4.10E-09 |
| 749 | 3605.H21.gz43_513961 | X63507 | M.musculus HOX-3.5 gene | 7.80E-05 |
| 750 | 3605.I19.gz43_513930 | AK002100 | Homo sapiens cDNA FLJ11238 fis, clone PLACE1008532 | 3.30E-11 |
| 751 | 3605.J16.gz43_513883 | AF039197 | Gallus gallus Pax-9 gene, putative 5' regulatory sequence | 1.00E-07 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 752 | 3605.K19.gz43_513932 | X63853 | S.cerevisiae MAT locus genes BUD5, mat-alpha1, mat-alpha2, YCR724 and YCR725 | 8.00E-06 |
| 753 | 3605.M17.gz43_513902 | M30931 | Simian immunodeficiency virus (SIV) proviral, complete genome | 3.70E-05 |
| 754 | 3605.N04.gz43_513695 | AF169388 | Mus musculus alpha 4 collagen IV (Col4a4) mRNA, complete cds | 8.90E-05 |
| 755 | 3605.N09.gz43_513775 | AF029111 | Adelius sp. 16S ribosomal RNA gene, mitochondrial gene for mitochondrial RNA, partial sequence | 2.80E-07 |
| 756 | 3605.N12.gz43_513823 | BC000358 | Homo sapiens, protein kinase, AMP-activated, gamma 1 non-catalytic subunit, clone MGC:8666 IMAGE:2964434, mRNA, complete cds | 3.90E-47 |
| 757 | 3605.N16.gz43_513887 | X95301 | D.rerio mRNA for HER-5 protein | 1.00E-06 |
| 758 | 3608.B06.gz43_514099 | X00004 | .taurus gene encoding pituitary glycoprotein hormone alpha subunit, exons 3 & 4 | 6.30E-08 |
| 759 | 3608.B12.gz43_514195 | X00525 | Mouse 28S ribosomal RNA | 3.10E-13 |
| 760 | 3608.B24.gz43_514387 | AF269848 | Staphylococcus epidermidis strain SR1 clone step.1026e06 genomic sequence | 2.00E-06 |
| 761 | 3608.C18.gz43_514292 | BC000387 | Homo sapiens, U6 snRNA-associated Sm-like protein, clone MGC:8433 IMAGE:2821171, mRNA, complete cds | 2.50E-10 |
| 762 | 3608.E17.gz43_514278 | BC008245 | Homo sapiens, clone IMAGE:3875012, mRNA | 1.00E-06 |
| 763 | 3608.E20.gz43_514326 | U86646 | Ailurus fulgens beta casein gene, exon 7, partial cds | 4.70E-07 |
| 764 | 3608.F13.gz43_514215 | AF125672 | Homo sapiens silencing mediator of retinoic acid and thyroid hormone receptor extended isoform (SMRTE) mRNA, complete cds | 2.00E-06 |
| 765 | 3608.G09.gz43_514152 | AE001066 | Archaeoglobus fulgidus section 41 of 172 of the complete genome | 4.00E-06 |
| 766 | 3608.H05.gz43_514089 | AJ224981 | Mus musculus calpain 3 gene, exon 1 | 3.00E-06 |
| 767 | 3608.H14.gz43_514233 | AE007394 | Streptococcus pneumoniae section 77 of 194 of the complete genome | 3.20E-05 |
| 768 | 3608.H18.gz43_514297 | Z36046 | S.cerevisiae chromosome II reading frame ORF YBR177c | 7.00E-06 |
| 769 | 3608.J17.gz43_514283 | AF024648 | Arabidopsis thaliana receptor-like serine/threonine kinase (RKF1) mRNA, complete cds | 8.00E-06 |
| 770 | 3608.J24.gz43_514395 | AJ002258 | Rattus Norvegicus mRNA for Prx3A protein | 3.60E-07 |
| 771 | 3608.K03.gz43_514060 | M83199 | Simmondsia chinensis stearoyl-acyl carrier protein desaturase mRNA, complete cds | 2.50E-07 |
| 772 | 3608.K14.gz43_514236 | AK026999 | Homo sapiens cDNA: FLJ23346 fis, clone HEP13716 | 2.00E-06 |
| 773 | 3608.L07.gz43_514125 | M32684 | Homo sapiens ITGB3 gene, intron 13, fragment B, partial sequence | 3.60E-07 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 774 | 3608.L14.gz43_514237 | Z34845 | H.sapiens serotonin transporter gene | 8.60E-07 |
| 775 | 3608.N09.gz43_514159 | AK022341 | Homo sapiens cDNA FLJ12279 fis, clone MAMMA1001743, weakly similar to Y BOX BINDING PROTEIN-1 | 2.00E-06 |
| 776 | 3608.N19.gz43_514319 | M15085 | T.brucei expressed copy of the ILTat 1.3 variable surface glycoprotein gene, 5' flank | 7.80E-08 |
| 777 | 3608.N20.gz43_514335 | AF026169 | Homo sapiens SALF (SALF) mRNA, complete cds | 1.00E-05 |
| 778 | 3608.O04.gz43_514080 | U85193 | Human nuclear factor I-B2 (NFIB2) mRNA, complete cds | 7.10E-07 |
| 779 | 3608.P22.gz43_514369 | AF124241 | Callerya australis chloroplast tRNA-Leu (trnL) gene, intron sequence | 3.90E-07 |
| 780 | 3611.A17.gz43_514658 | X01412 | Drosophila melanogaster genes for tRNA-Val and tRNA-Pro (90BC tRNA locus) | 2.00E-06 |
| 781 | 3611.B11.gz43_514563 | AL049938 | Homo sapiens mRNA; cDNA DKFZp564P1916 (from clone DKFZp564P1916); partial cds | 9.80E-10 |
| 782 | 3611.B16.gz43_514643 | M86514 | Rat proline-rich protein mRNA, 3' end | 1.30E-05 |
| 783 | 3611.C09.gz43_514532 | U55950 | Pleurodeles waltl cytochrome b (CYT-b) gene, mitochondrial gene encoding mitochondrial protein, partial cds | 2.00E-06 |
| 784 | 3611.E07.gz43_514502 | AF261009 | Lethrinus miniatus clone 89rte, microsatellite sequence | 1.70E-12 |
| 785 | 3611.E12.gz43_514582 | M60200 | Rat vitamin D binding protein gene, exons 5 and 6 | 1.50E-05 |
| 786 | 3611.E20.gz43_514710 | BC002458 | Homo sapiens, clone IMAGE:3343171, mRNA, partial cds | 2.00E-06 |
| 787 | 3611.F15.gz43_514631 | U28328 | Bos taurus dinucleotide repeat RM154, tandem repeat region | 4.30E-27 |
| 788 | 3611.H10.gz43_514553 | AE003147 | Drosophila melanogaster genomic scaffold 142000013385388, complete sequence | 6.00E-07 |
| 789 | 3611.H22.gz43_514745 | X16135 | Human mRNA for novel heterogeneous nuclear RNP protein, L protein | 7.00E-06 |
| 790 | 3611.I04.gz43_514458 | AK001460 | Homo sapiens cDNA FLJ10598 fis, clone NT2RP2004841 | 5.10E-44 |
| 791 | 3611.I13.gz43_514602 | M58380 | Arabidopsis thaliana peroxidase (neutral, prxCa) gene, complete cds | 3.00E-06 |
| 792 | 3611.J04.gz43_514459 | S81486 | p53 {alternatively spliced, intron 9} [human, Genomic Mutant, 133 nt] | 1.20E-07 |
| 793 | 3611.J15.gz43_514635 | AC008240 | Leishmania major chromosome 22 clone L9259 strain Friedlin, complete sequence | 4.90E-05 |
| 794 | 3611.J17.gz43_514667 | Z17425 | Lilium speciosum for two putative cds's | 8.90E-07 |
| 795 | 3611.J22.gz43_514747 | U60736 | Human IgHC locus intergenic sequence | 4.60E-07 |
| 796 | 3611.K01.gz43_514412 | AE001377 | Plasmodium falciparum chromosome 2, section 14 of 73 of the complete sequence | 3.00E-06 |
| 797 | 3611.K12.gz43_514588 | X02367 | Glaucoma chattoni rDNA 3' NTS | 9.80E-08 |
| 798 | 3611.L22.gz43_514749 | U19361 | Petromyzon marinus neurofilament subunit NF-180 mRNA, complete cds | 5.40E-08 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 799 | 3611.M18.gz43_514686 | X95301 | D.rerio mRNA for HER-5 protein | 1.00E-06 |
| 800 | 3611.M24.gz43_514782 | AF010239 | Caenorhabditis elegans glutathione S-transferase (CeGST1) mRNA, complete cds | 7.70E-07 |
| 801 | 3611.N01.gz43_514415 | L19300 | Staphylococcus aureus DNA sequence encoding three ORFs, complete cds; prophage phi-11 sequence homology, 5' flank | 1.00E-06 |
| 802 | 3611.N09.gz43_514543 | U50382 | Danio rerio beta and alpha globin genes, partial cds | 7.00E-06 |
| 803 | 3611.O16.gz43_514656 | AB056785 | Macaca fascicularis brain cDNA clone:QnpA-11655, full insert sequence | 6.60E-07 |
| 804 | 3611.P08.gz43_514529 | AK026905 | Homo sapiens cDNA: FLJ23252 fis, clone COL04668 | 8.00E-06 |
| 805 | 3614.C18.gz43_515060 | AF239178 | Paracoccidioides brasiliensis lon proteinase gene, complete cds; nuclear gene for mitochondrial product | 5.00E-06 |
| 806 | 3614.D14.gz43_514997 | AB017511 | Hydra magnipapillata mRNA for PLC-betaH1, complete cds | 1.20E-05 |
| 807 | 3614.D21.gz43_515109 | L10713 | Pig trinucleotide repeat | 1.80E-05 |
| 808 | 3614.E06.gz43_514870 | X99739 | M.musculus mRNA for UBC9 protein, containing ubiquitin box | 9.10E-07 |
| 809 | 3614.F22.gz43_515127 | AK021490 | Homo sapiens cDNA FLJ11428 fis, clone HEMBA1001071, highly similar to PROCOLLAGEN ALPHA 1(III) CHAIN PRECURSOR | 2.00E-06 |
| 810 | 3614.G20.gz43_515096 | M86514 | Rat proline-rich protein mRNA, 3' end | 1.30E-05 |
| 811 | 3614.H09.gz43_514921 | AF068289 | Homo sapiens HDCMD34P mRNA, complete cds | 6.60E-11 |
| 812 | 3614.H22.gz43_515129 | X62423 | P.falciparum pol delta gene for DNA polymerase delta | 4.00E-06 |
| 813 | 3614.J07.gz43_514891 | X81027 | H.sapiens tal-1 DNA | 1.30E-05 |
| 814 | 3614.K22.gz43_515132 | X63073 | Pseudanabaena sp. cpeBA operon encoding phycoerythrin beta and alpha subunits | 1.60E-05 |
| 815 | 3614.L13.gz43_514989 | V01561 | Mouse dispersed repetitive DNA sequences of the R-family and simple sequence DNA; member of the B1 family of mouse dispersed repetitive DNA sequences | 3.00E-06 |
| 816 | 3614.M08.gz43_514910 | AF272983 | Homo sapiens SRC tyrosine kinase gene, exons 1alpha and 1a, alternatively spliced | 4.00E-06 |
| 817 | 3614.O02.gz43_514816 | X58913 | Mitochondrion Drosophila eugracilis ND2 and COI genes (partial) and genes for tRNA-Trp, tRNA-Tyr, and tRNA-Cys | 8.50E-08 |
| 818 | 3614.O07.gz43_514896 | AL031538 | S.pombe chromosome III cosmid c1906 | 9.80E-07 |
| 819 | 3614.O16.gz43_515040 | AB056785 | Macaca fascicularis brain cDNA clone:QnpA-11655, full insert sequence | 2.00E-06 |
| 820 | 3614.P11.gz43_514961 | X91656 | M.musculus Srp20 gene | 4.60E-05 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 821 | 3614.P16.gz43_515041 | Z58907 | H.sapiens CpG island DNA genomic Msel fragment, clone 116a6, forward read cpg116a6.ft1a | 3.20E-70 |
| 822 | 3617.B16.gz43_515411 | AF098275 | Homo sapiens PSI2TOM20 pseudogene, complete sequence | 1.10E-67 |
| 823 | 3617.C21.gz43_515492 | AJ009913 | Bos taurus plp gene | 3.40E-05 |
| 824 | 3617.F10.gz43_515319 | L07487 | Bradyrhizobium japonicum heme-copper oxidase subunit I homolog (fixN), cytochrome c (fixO), transmembrane proteins (fixO and fixQ) diheme cytochrome c (fixP) and fixG genes, complete cds | 6.70E-05 |
| 825 | 3617.H16.gz43_515417 | X54192 | O.sativa GluB-2 gene for glutelin | 2.00E-06 |
| 826 | 3617.I01.gz43_515178 | AL513316 | Human DNA sequence from clone RP11-522O3 on chromosome 10, complete sequence [Homo sapiens] | 7.20E-08 |
| 827 | 3617.L16.gz43_515421 | AE007662 | Clostridium acetobutylicum ATCC824 section 150 of 356 of the complete genome | 3.00E-06 |
| 828 | 3617.L21.gz43_515501 | AL031538 | S.pombe chromosome III cosmid c1906 | 1.00E-06 |
| 829 | 3617.M08.gz43_515294 | X64802 | H.sapiens F8 mRNA for Interleukin-1-like species | 3.40E-58 |
| 830 | 3617.M13.gz43_515374 | Z79239 | H.sapiens flow-sorted chromosome 6 TaqI fragment, SC6pA26F6 | 1.10E-07 |
| 831 | 3617.N05.gz43_515247 | AF387666 | Mandrillus cytomegalovirus strain OCOM6-2 glycoprotein B (gB) gene, partial cds | 1.00E-06 |
| 832 | 3617.N10.gz43_515327 | AB017511 | Hydra magnipapillata mRNA for PLC-betaH1, complete cds | 1.10E-05 |
| 833 | 3617.N14.gz43_515391 | AJ249346 | Mus musculus Ankrd2 gene for ankyrin repeat domain 2 (stretch responsive muscle), exons 1-9 | 1.00E-05 |
| 834 | 3617.N19.gz43_515471 | U27037 | Fistulina hepatica mitochondrial small subunit ribosomal RNA, mitochondrial gene, partial sequence | 2.00E-06 |
| 835 | 3617.P11.gz43_515345 | AK002100 | Homo sapiens cDNA FLJ11238 fis, clone PLACE1008532 | 1.20E-13 |
| 836 | 3617.P12.gz43_515361 | U04860 | Rattus norvegicus Sprague-Dawley Ah receptor mRNA, complete cds | 8.00E-05 |
| 837 | 3617.P13.gz43_515377 | AE007356 | Streptococcus pneumoniae section 39 of 194 of the complete genome | 3.80E-05 |
| 838 | 3620.B03.gz43_515810 | AF238884 | Botrytis virus F, complete genome | 6.00E-06 |
| 839 | 3620.B24.gz43_516146 | AF244812 | Homo sapiens SCAN domain-containing protein 2 (SCAND2) gene, complete cds, alternatively spliced | 1.30E-07 |
| 840 | 3620.E12.gz43_515957 | X95301 | D.rerio mRNA for HER-5 protein | 1.00E-06 |
| 841 | 3620.E13.gz43_515973 | X52289 | Human (D21S167) DNA segment containing (GT)19 repeat | 2.50E-19 |
| 842 | 3620.E17.gz43_516037 | AJ002414 | Arabidosis thaliana mRNA for a hnRNP-like protein | 9.70E-08 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 843 | 3620.E19.gz43_516069 | X16982 | Drosophila melanogaster micropia-Dm11 3'flanking DNA | 2.70E-07 |
| 844 | 3620.E23.gz43_516133 | Z49438 | S.cerevisiae chromosome X reading frame ORF YJL163c | 3.00E-06 |
| 845 | 3620.E24.gz43_516149 | M75883 | Human sterol carrier protein X/sterol carrier protein 2 mRNA, complete cds | 8.00E-06 |
| 846 | 3620.G17.gz43_516039 | U92971 | Human protease-activated receptor 3 (PAR3) mRNA, complete cds | 3.80E-07 |
| 847 | 3620.G23.gz43_516135 | X66979 | X.laevis mRNA XLFLI | 1.60E-05 |
| 848 | 3620.J18.gz43_516058 | U37373 | Xenopus laevis tail-specific thyroid hormone up-regulated (gene 5) mRNA, complete cds | 3.00E-06 |
| 849 | 3620.K19.gz43_516075 | U31780 | Human papillomavirus type 22, complete genome | 5.00E-06 |
| 850 | 3620.K24.gz43_516155 | M95627 | Homo sapiens angio-associated migratory cell protein (AAMP) mRNA, complete cds | 6.00E-06 |
| 851 | 3620.O23.gz43_516143 | L11172 | Plasmodium falciparum RNA polymerase I gene, complete cds | 1.00E-05 |
| 852 | 3623.B07.gz43_516258 | AF132745 | Mus musculus Sox2 gene, regulatory region sequence | 7.70E-07 |
| 853 | 3623.E03.gz43_516197 | X82566 | M.musculus glyT1 gene (exon 0a) | 1.80E-09 |
| 854 | 3623.E15.gz43_516389 | AF104420 | Porcine transmissible gastroenteritis virus RNA dependent RNA polymerase gene, partial cds; virus envelope protein spike (S), envelope protein (sM), envelope protein (M), and nucleoprotein (N) genes, complete cds; and unknown genes | 2.90E-05 |
| 855 | 3623.F03.gz43_516198 | AJ009936 | Homo sapiens mRNA for nuclear hormone receptor PRR1 | 1.70E-05 |
| 856 | 3623.F20.gz43_516470 | U22657 | Mus musculus genomic locus related to cellular morphology | 5.80E-05 |
| 857 | 3623.G14.gz43_516375 | AB035309 | Paramecium caudatum PcTERT mRNA for telomerase reverse transcriptase, complete cds | 3.00E-06 |
| 858 | 3623.H07.gz43_516264 | Z17324 | Homo sapiens of MUC1 gene encoding Mucin | 1.80E-07 |
| 859 | 3623.H10.gz43_516312 | AB033070 | Homo sapiens mRNA for KIAA1244 protein, partial cds | 2.80E-05 |
| 860 | 3623.H23.gz43_516520 | AF131763 | Homo sapiens clone 25232 mRNA sequence | 1.70E-05 |
| 861 | 3623.I08.gz43_516281 | M60421 | Human cytochrome P450scc gene, 5' end and promoter region | 2.80E-05 |
| 862 | 3623.I11.gz43_516329 | AK013191 | Mus musculus 10, 11 days embryo cDNA, RIKEN full-length enriched library, clone:2810429I04, full insert sequence | 3.00E-06 |
| 863 | 3623.L05.gz43_516236 | AJ131991 | Linum usitatissimum target sequence for LIS-1 insertion in Pl | 3.00E-06 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 864 | 3623.L24.gz43_516540 | U09377 | Arabidopsis thaliana GF14chi isoform (GRF1) gene, complete cds | 3.00E-06 |
| 865 | 3623.M10.gz43_516317 | AF071743 | Homo sapiens topoisomerase II alpha (TOP2A) gene, exons 25, 26, and 27 | 4.00E-06 |
| 866 | 3623.N23.gz43_516526 | U57489 | Eubacterium sp. VPI 12708 bile acid-inducible operon bile acid-coenzyme A ligase (baiB), BaiC, BaiD, bile acid 7-alpha dehydratase (baiE), 3-alpha hydroxysteroid dehydrogenase (baiA2), BaiF, bile acid transporter (baiG), NADH:flavin oxidoreductase (bai> | 3.70E-05 |
| 867 | 3623.P22.gz43_516512 | U37761 | Human H1 histamine receptor gene, 5'-flanking region | 1.40E-12 |
| 868 | 3626.A10.gz43_516689 | D30745 | Xenopus laevis MRP RNA gene | 2.00E-07 |
| 869 | 3626.C16.gz43_516787 | AF241271 | Bos taurus ZFY gene, intron | 1.60E-08 |
| 870 | 3626.E07.gz43_516645 | AF053496 | Caenorhabditis elegans beta chain spectrin homolog Sma1 (sma1) mRNA, complete cds | 2.00E-06 |
| 871 | 3626.F03.gz43_516582 | AJ009771 | Homo sapiens mRNA for putative RING finger protein, partial | 2.00E-06 |
| 872 | 3626.G01.gz43_516551 | BC010926 | Homo sapiens, Similar to H4 histone family, member A, clone MGC:13512 IMAGE:4273904, mRNA, complete cds | 1.00E-43 |
| 873 | 3626.I20.gz43_516857 | AK025762 | Homo sapiens cDNA: FLJ22109 fis, clone HEP18091 | 5.80E-07 |
| 874 | 3626.I23.gz43_516905 | S55615 | (156)=G surface antigen {3' region, restriction fragment EG4} [Paramecium primaurelia, Genomic, 407 nt] | 3.40E-07 |
| 875 | 3626.M13.gz43_516749 | AE001398 | Plasmodium falciparum chromosome 2, section 35 of 73 of the complete sequence | 4.00E-06 |
| 876 | 3626.M15.gz43_516781 | AF090925 | Homo sapiens clone HQ0452 PRO0452 mRNA, partial cds | 3.10E-07 |
| 877 | 3626.N07.gz43_516654 | Z58907 | H.sapiens CpG island DNA genomic Mse1 fragment, clone 116a6, forward read cpg116a6.ft1a | 2.90E-70 |
| 878 | 3626.N24.gz43_516926 | AF041373 | Rattus norvegicus clathrin assembly protein short form (CALM) mRNA, complete cds | 8.90E-08 |
| 879 | 3626.O08.gz43_516671 | D10445 | Mouse mRNA for protein C, complete cds | 5.00E-06 |
| 880 | 3626.P11.gz43_516720 | L48479 | Homo sapiens (subclone 6_h1 from P1 H21) DNA sequence | 2.20E-07 |
| 881 | 3626.P14.gz43_516768 | X15028 | Chicken hsp90 gene for 90 kDa-heat shock protein 5'-end | 3.80E-05 |
| 882 | 3629.A16.gz43_517169 | U16958 | Mus musculus pre-T cell receptor alpha-type chain precursor mRNA, complete cds | 4.00E-06 |
| 883 | 3629.B14.gz43_517138 | X16982 | Drosophila melanogaster micropia-Dm11 3'flanking DNA | 2.50E-07 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 884 | 3629.C14.gz43_517139 | Z22537 | C.parvum precursor of oocyst wall protein | 5.00E-06 |
| 885 | 3629.E01.gz43_516933 | D00621 | Sus scrofa gene for follicle stimulation hormone beta subunit, exons 1, 2, 3, complete cds | 3.50E-05 |
| 886 | 3629.E20.gz43_517237 | AE006900 | Sulfolobus solfataricus section 259 of 272 of the complete genome | 9.00E-06 |
| 887 | 3629.F24.gz43_517302 | Y10531 | Clostridium perfringens sod gene for superoxide dismutase | 2.00E-06 |
| 888 | 3629.H10.gz43_517080 | J03654 | Human immunodeficiency virus type 2, isolate HIV2FG | 8.00E-06 |
| 889 | 3629.H12.gz43_517112 | AF017266 | Danio rerio glutamate decarboxylase (GAD67) mRNA, partial cds | 6.50E-07 |
| 890 | 3629.I11.gz43_517097 | AF020810 | Salmonella enterica VirK (virK), Mig-14 (mig-14), NxiA (nxiA), TctE (tctE), TctD (tctD), TctC (tctC), TctB (tctB), and TctA (tctA) genes, complete cds; and O360 (o360) gene, partial cds | 3.00E-06 |
| 891 | 3629.I16.gz43_517177 | AE007643 | Clostridium acetobutylicum ATCC824 section 131 of 356 of the complete genome | 4.40E-05 |
| 892 | 3629.J03.gz43_516970 | AB017511 | Hydra magnipapillata mRNA for PLC-betaH1, complete cds | 1.10E-05 |
| 893 | 3629.J07.gz43_517034 | M20782 | Human alpha-2-plasmin inhibitor gene, exons 2 to 5 | 2.90E-11 |
| 894 | 3632.C11.gz43_517475 | AF026148 | Perilla frutescens beta-ketoacyl-ACP synthase I (KAS I) mRNA, complete cds | 1.00E-06 |
| 895 | 3632.C17.gz43_517571 | U50534 | Human BRCA2 region, mRNA sequence CG003 | 1.00E-05 |
| 896 | 3632.F07.gz43_517414 | M12036 | Human tyrosine kinase-type receptor (HER2) gene, partial cds | 4.70E-10 |
| 897 | 3632.G01.gz43_517319 | AC006621 | Caenorhabditis elegans cosmid C52A10, complete sequence | 3.40E-05 |
| 898 | 3632.I20.gz43_517625 | AK024381 | Homo sapiens cDNA FLJ14319 fis, clone PLACE3000406 | 9.00E-06 |
| 899 | 3632.K20.gz43_517627 | M27634 | Vaccinia virus P4a major core protein gene, complete cds | 9.60E-05 |
| 900 | 3632.M08.gz43_517437 | X75304 | H.sapiens giantin mRNA | 8.00E-06 |
| 901 | 3632.M13.gz43_517517 | U18191 | Human HLA class I genomic survey sequence | 3.20E-07 |
| 902 | 3632.M19.gz43_517613 | AF012131 | Homo sapiens brachyury variant B (TBX1) mRNA, complete cds | 3.70E-07 |
| 903 | 3632.N13.gz43_517518 | AF287491 | Oncorhynchus mykiss MHC class I heavy chain precursor (Onmy-UBA) mRNA, Onmy-UBA*601 allele, complete cds | 2.00E-06 |
| 904 | 3632.N21.gz43_517646 | X62423 | P.falciparum pol delta gene for DNA polymerase delta | 4.00E-06 |
| 905 | 3632.O06.gz43_517407 | BC009868 | Homo sapiens, replication protein A3 (14kD), clone MGC:16404 IMAGE:3940438, mRNA, complete cds | 1.40E-18 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 906 | 3632.P07.gz43_517424 | AE001066 | Archaeoglobus fulgidus section 41 of 172 of the complete genome | 3.00E-06 |
| 907 | 3635.A06.gz43_517777 | AK005546 | Mus musculus adult female placenta cDNA, RIKEN full-length enriched library, clone:1600027G01; full insert sequence | 1.40E-07 |
| 908 | 3635.A08.gz43_517809 | Z49280 | S.cerevisiae chromosome X reading frame ORF YJL005w | 6.00E-06 |
| 909 | 3635.A13.gz43_517889 | AF143236 | Homo sapiens apoptosis related protein APR 2 mRNA, complete cds | 2.00E-06 |
| 910 | 3635.D07.gz43_517796 | M58150 | Bovine lactoperoxidase (LPO) mRNA, complete cds | 3.10E-05 |
| 911 | 3635.F01.gz43_517702 | Y19128 | Homo sapiens enteropeptidase gene, exon 6 | 3.00E-09 |
| 912 | 3635.F06.gz43_517782 | X63073 | Pseudanabaena sp. cpeBA operon encoding phycoerythrin beta and alpha subunits | 1.50E-05 |
| 913 | 3635.F10.gz43_517846 | AF107688 | Aedes aegypti clone 431 Feilai family of SINES | 3.50E-05 |
| 914 | 3635.H20.gz43_518008 | AE000613 | Helicobacter pylori 26695 section 91 of 134 of the complete genome | 1.10E-05 |
| 915 | 3635.J06.gz43_517786 | U15018 | Dugbe virus L protein gene, complete cds | 1.10E-05 |
| 916 | 3635.J09.gz43_517834 | X85444 | G.pallida repetitive DNA element | 2.10E-08 |
| 917 | 3635.K05.gz43_517771 | AF090432 | Danio rerio serrateB mRNA, complete cds | 4.00E-06 |
| 918 | 3635.K06.gz43_517787 | AJ276631 | Capsicum annuum partial kn gene for Knolle protein, promoter region | 6.10E-07 |
| 919 | 3635.M18.gz43_517981 | AL591498 | Human DNA sequence from clone RP11-113L12 on chromosome 13, complete sequence [Homo sapiens] | 1.40E-05 |
| 920 | 3635.O01.gz43_517711 | AF081788 | Homo sapiens putative spliceosome associated protein mRNA, complete cds | 3.70E-30 |
| 921 | 3635.O14.gz43_517919 | X72224 | S.cerevisiae genes HSS1, NPL4 and HSP | 6.00E-06 |
| 922 | 3635.P17.gz43_517968 | AF242307 | Euphorbia esula sucrose transport protein mRNA, complete cds | 2.90E-10 |
| 923 | 3635.P18.gz43_517984 | AF078780 | Caenorhabditis elegans cosmid C04F2, complete sequence | 1.74E-04 |
| 924 | 3638.A02.gz43_518097 | M17988 | Spiroplasma virus 4 (SpV4) replicative form, complete genome | 4.00E-06 |
| 925 | 3638.A24.gz43_518449 | AF064079 | Plasmodium gallinaceum endochitinase precursor, mRNA, complete cds | 1.60E-07 |
| 926 | 3638.F15.gz43_518310 | AJ297538 | Homo sapiens partial RARA gene, intron 2 | 4.00E-06 |
| 927 | 3638.H07.gz43_518184 | AK026258 | Homo sapiens cDNA: FLJ22605 fis, clone HSI04743 | 2.00E-06 |
| 928 | 3638.J09.gz43_518218 | U89651 | Homo sapiens matrix metalloproteinase MMP Rasi-1 gene, promoter region | 8.10E-08 |
| 929 | 3638.K06.gz43_518171 | AL139329 | Human DNA sequence from clone RP11-228P1 on chromosome 6, complete sequence [Homo sapiens] | 4.40E-11 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 930 | 3638.L10.gz43_518236 | D26532 | Mouse mRNA for transcription factor PEBP2aB2, complete cds | 2.00E-08 |
| 931 | 3638.N05.gz43_518158 | X62294 | B.taurus mRNA for adrenal angiotensin II type-1 receptor | 9.00E-06 |
| 932 | 3643.D21.gz43_518788 | U17010 | Allomyces macrogynus mitochondrion NADH dehydrogenase subunit 5 (nad5) gene, complete cds | 1.80E-05 |
| 933 | 3643.E24.gz43_518837 | AL022342 | Human DNA sequence from clone RP1-29M10 on chromosome 20, complete sequence [Homo sapiens] | 6.70E-05 |
| 934 | 3643.F07.gz43_518566 | M73962 | Bovine pregnancy-associated glycoprotein 1 mRNA, complete cds | 6.00E-06 |
| 935 | 3643.G20.gz43_518775 | AF191214 | Homo sapiens isovaleryl dehydrogenase (IVD) gene, exons 1-3 | 1.00E-05 |
| 936 | 3643.G24.gz43_518839 | AK025682 | Homo sapiens cDNA: FLJ22029 fis, clone HEP08661 | 6.00E-06 |
| 937 | 3643.H09.gz43_518600 | AK024381 | Homo sapiens cDNA FLJ14319 fis, clone PLACE3000406 | 1.70E-05 |
| 938 | 3643.I01.gz43_518473 | AF000306 | Brassica napus steroid sulfotransferase 2 gene, complete cds | 3.00E-06 |
| 939 | 3643.I02.gz43_518489 | X58433 | B.subtillis cad gene for lysine decarboxylase | 2.30E-05 |
| 940 | 3643.I18.gz43_518745 | M14872 | Mouse GnRH-GAP gene encoding gonadotropin-releasing hormone and Gn-RH-associated peptide (GAP) | 4.00E-06 |
| 941 | 3643.I24.gz43_518841 | BC003813 | Mus musculus, clone MGC:6139 IMAGE:3487295, mRNA, complete cds | 2.30E-07 |
| 942 | 3643.K06.gz43_518555 | AL050124 | Homo sapiens mRNA; cDNA DKFZp586E151 (from clone DKFZp586E151) | 1.60E-07 |
| 943 | 3643.L01.gz43_518476 | AJ278429 | Mus musculus partial Prkar1a gene for cAMP-dependent protein kinase regulatory subunit RIalpha, exons 8-10 and 3'UTR | 3.00E-06 |
| 944 | 3643.N24.gz43_518846 | BC006511 | Homo sapiens, clone IMAGE:3010441, mRNA | 1.00E-05 |
| 945 | 3643.O16.gz43_518719 | AE002303 | Chlamydia muridarum, section 34 of 85 of the complete genome | 1.10E-05 |
| 946 | 3643.O18.gz43_518751 | V00248 | Drosophila gene for yolk protein I (vitellogenin) | 2.00E-06 |
| 947 | 3643.O21.gz43_518799 | AE000614 | Helicobacter pylori 26695 section 92 of 134 of the complete genome | 1.40E-05 |
| 948 | 3643.P13.gz43_518672 | Y17693 | Bungarus multicinctus gene encoding alpha-bungarotoxin, V31 variant | 2.00E-07 |
| 949 | 3643.P14.gz43_518688 | AF109352 | Euperipatoides rowelli microsatellite P18 sequence | 8.80E-10 |
| 950 | 3646.A07.gz43_518945 | X55137 | H. giganteus type II restriction-modification system HgiBI | 3.00E-06 |
| 951 | 3646.A09.gz43_518977 | AF074963 | Rattus norvegicus endothelin-B receptor (EDNRB) gene, partial cds | 2.10E-07 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 952 | 3646.A12.gz43_519025 | AF176208 | Homo sapiens EcoRI-HindIII fragment upstream of exon 1 of the c-myc gene | 1.60E-05 |
| 953 | 3646.A13.gz43_519041 | X89445 | O.chalybea DNA for narB gene and partial ORFs | 4.00E-05 |
| 954 | 3646.B20.gz43_519154 | M86514 | Rat proline-rich protein mRNA, 3' end | 1.60E-05 |
| 955 | 3646.C06.gz43_518931 | Z71180 | Caenorhabditis elegans cosmid F22E12, complete sequence | 2.03E-04 |
| 956 | 3646.C16.gz43_519091 | U73608 | Hepatitis B virus, genome 7648 with G->A hypermutations | 2.30E-05 |
| 957 | 3646.E02.gz43_518869 | U11683 | Trypanoplasma borreli Tt-JH mitochondrion cytochrome c oxidase subunit 1 (cox1) gene, complete cds | 8.10E-07 |
| 958 | 3646.E20.gz43_519157 | AE006216 | Pasteurella multocida PM70 section 183 of 204 of the complete genome | 2.30E-05 |
| 959 | 3646.H04.gz43_518904 | AF043740 | Branchiostoma floridae amphioxus Otx transcription factor (Otx) mRNA, complete cds | 2.00E-06 |
| 960 | 3646.H09.gz43_518984 | AP000145 | Homo sapiens genomic DNA, chromosome 21q21.2, LL56-APP region, clone B2291C14-R44F3, segment 10/10, complete sequence | 2.90E-40 |
| 961 | 3646.H16.gz43_519096 | U22342 | Bacteriophage T270 integrase (int) gene, complete cds | 1.00E-07 |
| 962 | 3646.I01.gz43_518857 | X54486 | Human gene for C1-inhibitor | 6.80E-05 |
| 963 | 3646.J03.gz43_518890 | AB055372 | Macaca fascicularis brain cDNA, clone:QflA-12842 | 5.40E-190 |
| 964 | 3646.J22.gz43_519194 | AL133032 | Homo sapiens mRNA; cDNA DKFZp586B0317 (from clone DKFZp586B0317) | 2.00E-06 |
| 965 | 3646.K14.gz43_519067 | AF239178 | Paracoccidioides brasiliensis lon proteinase gene, complete cds; nuclear gene for mitochondrial product | 4.00E-06 |
| 966 | 3646.L17.gz43_519116 | Z58907 | H.sapiens CpG island DNA genomic Mse1 fragment, clone 116a6, forward read cpg116a6.ft1a | 2.50E-70 |
| 967 | 3646.O13.gz43_519055 | AL050391 | Homo sapiens mRNA; cDNA DKFZp586A181 (from clone DKFZp586A181); partial cds | 5.20E-08 |
| 968 | 3646.O16.gz43_519103 | X00331 | Drosophila virilis simple DNA sequence (pDV-161) | 5.20E-08 |
| 969 | 3646.P09.gz43_518992 | U04527 | Borrelia burgdorferi 212 DNA gyrase b subunit (gyrB) and ribonuclease P protein component (rnpA) genes, partial cds, DnaA protein (dnaA), DNA polymerase III beta subunit (dnaN), and ribosomal protein L34 (rpmH) genes, complete cds | 5.00E-06 |
| 970 | 3646.P14.gz43_519072 | AY032863 | Mus musculus chloride-formate exchanger mRNA, complete cds | 8.00E-06 |
| 971 | 3646.P17.gz43_519120 | U19569 | Human squamous cell carcinoma antigen (SCCA2) gene, exon 1 | 1.20E-07 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 972 | 3661.A08.gz43_519483 | AB017511 | Hydra magnipapillata mRNA for PLC-betaH1, complete cds | 1.20E-05 |
| 973 | 3661.D17.gz43_519630 | J03488 | Reovirus type 3 L2 gene encoding guanylyltransferase, complete cds | 3.00E-06 |
| 974 | 3661.D18.gz43_519646 | AB033024 | Homo sapiens mRNA for KIAA1198 protein, partial cds | 1.90E-11 |
| 975 | 3661.E19.gz43_519663 | AB014084 | Homo sapiens genomic DNA, chromosome 6p21.3, HLA class I region, Cosmid clone:TY7A5, complete sequence | 6.00E-05 |
| 976 | 3661.E23.gz43_519727 | AE001032 | Archaeoglobus fulgidus section 75 of 172 of the complete genome | 5.30E-05 |
| 977 | 3661.F14.gz43_519584 | X15063 | Plasmodium falciparum mRNA for major merozoite surface antigen gp195 | 6.80E-05 |
| 978 | 3661.G16.gz43_519617 | AF255609 | Homo sapiens high mobility group protein HMG1 gene, exons 1 and 2, partial cds | 2.70E-07 |
| 979 | 3661.G20.gz43_519681 | AK021558 | Homo sapiens cDNA FLJ11496 fis, clone HEMBA1001964 | 6.40E-09 |
| 980 | 3661.H11.gz43_519538 | Z30705 | Puumala virus (Evo/15Cg/93) gene for N protein | 3.90E-07 |
| 981 | 3661.H24.gz43_519746 | X66979 | X.laevis mRNA XLFLI | 1.60E-05 |
| 982 | 3661.I22.gz43_519715 | AF029887 | Caenorhabditis elegans UNC-129 (unc-129) mRNA, complete cds | 5.00E-06 |
| 983 | 3661.J15.gz43_519604 | AJ297538 | Homo sapiens partial RARA gene, intron 2 | 4.00E-06 |
| 984 | 3661.K22.gz43_519717 | AK002100 | Homo sapiens cDNA FLJ11238 fis, clone PLACE1008532 | 1.30E-13 |
| 985 | 3661.L19.gz43_519670 | AL589643 | Human DNA sequence from clone RP11-344C1 on chromosome 6, complete sequence [Homo sapiens] | 2.20E-05 |
| 986 | 3661.M03.gz43_519415 | Z57613 | H.sapiens CpG island DNA genomic Mse1 fragment, clone 187a12, forward read cpg187a12.ft1a | 1.20E-08 |
| 987 | 3661.M23.gz43_519735 | X79547 | Equus caballus mitochondrial DNA complete sequence | 5.80E-05 |
| 988 | 3661.P22.gz43_519722 | AF055668 | Mus musculus apoptosis-linked gene 4, deltaC form (Alg-4) mRNA, partial cds | 8.00E-06 |
| 989 | 3662.A13.gz43_519947 | Z49438 | S.cerevisiae chromosome X reading frame ORF YJL163c | 3.00E-06 |
| 990 | 3662.B13.gz43_519948 | AB045237 | Xenopus laevis XRPTPb mRNA for receptor-type protein tyrosine phosphatase beta.11, complete cds | 7.00E-06 |
| 991 | 3662.C10.gz43_519901 | BC007905 | Homo sapiens, Similar to retinal degeneration B beta, clone MGC:14375 IMAGE:4299595, mRNA, complete cds | 1.20E-09 |
| 992 | 3662.C15.gz43_519981 | M33864 | Human (cline HGL-3) interstitial retinoid-binding protein 3 (RBP3) gene, exon 1 | 1.20E-05 |
| 993 | 3662.F13.gz43_519952 | AB040935 | Homo sapiens mRNA for KIAA1502 protein, partial cds | 1.20E-61 |
| 994 | 3662.H14.gz43_519970 | AB032757 | Mus musculus gad65 gene for glutamate decarboxylase 65, partial cds | 8.00E-07 |

**Table 8**

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 995 | 3662.H23.gz43_520114 | AK013013 | Mus musculus 10, 11 days embryo cDNA, RIKEN full-length enriched library, clone:2810406L04, full insert sequence | 2.00E-06 |
| 996 | 3662.H24.gz43_520130 | D45371 | Human apM1 mRNA for GS3109 (novel adipose specific collagen-like factor), complete cds | 9.60E-10 |
| 997 | 3662.J05.gz43_519828 | M83554 | H.sapiens lymphocyte activation antigen CD30 mRNA, complete cds | 1.40E-05 |
| 998 | 3662.J08.gz43_519876 | Z11876 | B.hermsii vmp7 gene encoding Vmp7 outer membrane lipoprotein | 1.11E-04 |
| 999 | 3662.J09.gz43_519892 | AB011101 | Homo sapiens mRNA for KIAA0529 protein, partial cds | 6.30E-05 |
| 1000 | 3662.J16.gz43_520004 | U00484 | Anabaena PCC7120 protein kinase PknA (pknA) gene, complete cds | 2.00E-06 |
| 1001 | 3662.K03.gz43_519797 | AL390145 | Homo sapiens mRNA; cDNA DKFZp762C115 (from clone DKFZp762C115) | 1.40E-05 |
| 1002 | 3662.L05.gz43_519830 | U63635 | Schizosaccharomyces pombe RNA lariat debranching enzyme (Sp-dbr1) gene, complete cds | 5.80E-10 |
| 1003 | 3662.N24.gz43_520136 | Z30709 | L.helveticus genes for prolinase and putative ABC transporter | 3.70E-05 |
| 1004 | 3662.O02.gz43_519785 | AF084460 | Gallus gallus potassium channel Shaker alpha subunit variant cKv1.4(m) mRNA, complete cds | 6.90E-05 |
| 1005 | 3662.P03.gz43_519802 | AJ011456 | Schinziella tetragona matK gene (corresponding location in Tobacco: 963-1244) | 7.20E-08 |
| 1006 | 3663.A09.gz43_520267 | Z69608 | A.rara SSU rRNA gene (partial) | 3.30E-07 |
| 1007 | 3663.C08.gz43_520253 | Z50756 | Caenorhabditis elegans cosmid T08D10, complete sequence | 7.60E-07 |
| 1008 | 3663.C19.gz43_520429 | Z22672 | H.sapiens cacnl1a3 gene encoding skeletal muscle dhp-receptor alpha 1 subunit | 2.80E-07 |
| 1009 | 3663.E04.gz43_520191 | U89318 | Homo sapiens nucleophosmin phosphoprotein (NPM) gene, intron 9, partial sequence | 2.60E-07 |
| 1010 | 3663.F15.gz43_520368 | U66073 | Tritrichomonas foetus putative superoxide dismutase 1 (SOD1) gene, complete cds | 9.20E-07 |
| 1011 | 3663.F22.gz43_520480 | U36786 | Rattus norvegicus putative pheromone receptor VN7 mRNA, complete cds | 7.10E-07 |
| 1012 | 3663.G01.gz43_520145 | AK024359 | Homo sapiens cDNA FLJ14297 fis, clone PLACE1008941 | 9.50E-36 |
| 1013 | 3663.G08.gz43_520257 | L19339 | Molgula oculata zinc finger protein (manx) mRNA, complete cds | 5.20E-07 |
| 1014 | 3663.H20.gz43_520450 | X61307 | Staphylococcus aureus spa gene for protein A | 5.00E-06 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 1015 | 3663.J06.gz43_520228 | AE007916 | Agrobacterium tumefaciens strain C58 plasmid AT, section 44 of 50 of the complete sequence | 2.02E-04 |
| 1016 | 3663.J16.gz43_520388 | U38181 | Leuconostoc mesenteroides dextransucrase gene, complete cds | 3.90E-07 |
| 1017 | 3663.K02.gz43_520165 | X68339 | Mycoplasma-like organism (substrain ASHY) DNA for 16S rRNA | 5.00E-06 |
| 1018 | 3663.K13.gz43_520341 | AF155221 | Mus musculus matrix metalloproteinase 19 (Mmp19) mRNA, complete cds | 2.00E-06 |
| 1019 | 3663.L18.gz43_520422 | AB031056 | Solobacterium moorei gene for 16S rRNA, isolate:RCA59-74 | 1.00E-06 |
| 1020 | 3663.L24.gz43_520518 | D10445 | Mouse mRNA for protein C, complete cds | 6.00E-06 |
| 1021 | 3663.M24.gz43_520519 | AE001196 | Treponema pallidum section 12 of 87 of the complete genome | 5.20E-05 |
| 1022 | 3663.N09.gz43_520280 | AF081788 | Homo sapiens putative spliceosome associated protein mRNA, complete cds | 4.00E-20 |
| 1023 | 3663.N10.gz43_520296 | X62423 | P.falciparum pol delta gene for DNA polymerase delta | 4.00E-06 |
| 1024 | 3663.N12.gz43_520328 | AF178079 | Zygosaccharomyces rouxii ketoreductase (krd) mRNA, complete cds | 5.00E-06 |
| 1025 | 3663.N16.gz43_520392 | U41060 | Homo sapiens estrogen regulated LIV-1 protein (LIV-1) mRNA, complete cds | 2.00E-06 |
| 1026 | 3663.O07.gz43_520249 | D00442 | Grapevine fanleaf virus satellite RNA (RNA3), complete cds | 1.50E-08 |
| 1027 | 3663.O09.gz43_520281 | AK002141 | Homo sapiens cDNA FLJ11279 fis, clone PLACE1009444, highly similar to PHOSPHATIDYLINOSITOL 4-KINASE ALPHA (EC 2.7.1.67) | 5.30E-10 |
| 1028 | 3664.A11.gz43_520683 | U67525 | Methanococcus jannaschii section 67 of 150 of the complete genome | 4.00E-06 |
| 1029 | 3664.C21.gz43_520845 | AF064773 | Staphylococcus aureus extracellular enterotoxin type G precursor (SEG) gene, complete cds | 1.30E-07 |
| 1030 | 3664.D06.gz43_520606 | AF178079 | Zygosaccharomyces rouxii ketoreductase (krd) mRNA, complete cds | 5.00E-06 |
| 1031 | 3664.D12.gz43_520702 | U10519 | Human DNA polymerase beta gene, exon 5 | 2.00E-07 |
| 1032 | 3664.D17.gz43_520782 | AK027226 | Homo sapiens cDNA: FLJ23573 fis, clone LNG12520 | 4.90E-07 |
| 1033 | 3664.E18.gz43_520799 | AF317204 | Mus musculus C-type lectin superfamily 1 gene, complete cds | 3.20E-05 |
| 1034 | 3664.E23.gz43_520879 | AB050903 | Mus musculus mRNA for a4 subunit isoform, complete cds | 3.00E-06 |
| 1035 | 3664.E24.gz43_520895 | Z92793 | Caenorhabditis elegans cosmid H15M21, complete sequence | 1.20E-05 |
| 1036 | 3664.G12.gz43_520705 | AF211482 | Dictyostelium discoideum SdhA (sdhA) gene, complete cds | 2.30E-09 |
| 1037 | 3664.G20.gz43_520833 | M14450 | Rat thyrotropin (TSH) beta-subunit gene, exons 2 and 3 | 4.00E-06 |
| 1038 | 3664.H15.gz43_520754 | Y11270 | E.histolytica INO1 gene | 2.00E-06 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 1039 | 3664.H22.gz43_520866 | X97773 | B.taurus mRNA for mitochondrial tricarboxylate carrier protein | 1.20E-05 |
| 1040 | 3664.J12.gz43_520708 | M58150 | Bovine lactoperoxidase (LPO) mRNA, complete cds | 3.20E-05 |
| 1041 | 3664.J23.gz43_520884 | U67463 | Methanococcus jannaschii section 5 of 150 of the complete genome | 3.00E-06 |
| 1042 | 3664.K16.gz43_520773 | Z83118 | Caenorhabditis elegans cosmid M04D5, complete sequence | 2.70E-07 |
| 1043 | 3664.K19.gz43_520821 | U36927 | Plasmodium yoelii rhoptry protein gene, complete cds | 3.00E-05 |
| 1044 | 3664.L21.gz43_520854 | AF057695 | Haemophilus ducreyi strain 35000 putative phosphomannomutase (pmm) gene, partial cds; large supernatant protein 1 (lspA1) gene, complete cds; and putative GMP synthase (guaA) gene, partial cds | 2.15E-04 |
| 1045 | 3664.O22.gz43_520873 | U43574 | Hydra vulgaris nucleoporin p62 gene, complete cds | 7.00E-06 |
| 1046 | 3664.P12.gz43_520714 | AF030883 | Mus musculus tRNA-His gene, complete sequence; platelet-activating factor acetylhydrolase Ib alpha subunit (Pafaha-ps1) pseudogene, complete sequence; and tRNA-Glu gene, complete sequence | 9.00E-06 |
| 1047 | 3664.P18.gz43_520810 | Z47735 | H.sapiens NFKB1 gene, exons 11 & 12 | 1.32E-04 |
| 1048 | 3665.A23.gz43_521259 | X66979 | X.laevis mRNA XLFLI | 1.60E-05 |
| 1049 | 3665.B01.gz43_520908 | M90058 | Human serglycin gene, exons 1,2, and 3 | 4.00E-06 |
| 1050 | 3665.B12.gz43_521084 | AK020877 | Mus musculus adult retina cDNA, RIKEN full-length enriched library, clone:A930019H03, full insert sequence | 7.10E-07 |
| 1051 | 3665.E11.gz43_521071 | AB024030 | Arabidopsis thaliana genomic DNA, chromosome 5, TAC clone:K5A21 | 9.00E-06 |
| 1052 | 3665.E20.gz43_521215 | X76584 | H.sapiens simple DNA sequence region clone wg1h1 | 6.80E-08 |
| 1053 | 3665.H20.gz43_521218 | X95301 | D.rerio mRNA for HER-5 protein | 9.50E-07 |
| 1054 | 3665.K01.gz43_520917 | X04653 | Mouse mRNA for Ly-6 alloantigen (Ly-6E.1) | 1.30E-05 |
| 1055 | 3665.M01.gz43_520919 | AF098352 | Wiseana copularis haplotype southern cytochrome oxidase subunit I and cytochrome oxidase subunit II genes, partial cds; mitochondrial genes for mitochondrial products | 5.80E-07 |
| 1056 | 3665.M21.gz43_521239 | AF257480 | Rana temporaria microsatellite SB80 sequence | 3.30E-09 |
| 1057 | 3665.M23.gz43_521271 | Y10623 | C.pallidivittatus globin gene cluster E | 1.10E-05 |
| 1058 | 3665.N24.gz43_521288 | X95301 | D.rerio mRNA for HER-5 protein | 1.00E-06 |
| 1059 | 3665.O06.gz43_521001 | AE007033 | Mycobacterium tuberculosis CDC1551, section 119 of 280 of the complete genome | 7.40E-05 |
| 1060 | 3665.O14.gz43_521129 | AB033094 | Homo sapiens mRNA for KIAA1268 protein, partial cds | 2.10E-08 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 1061 | 3665.O15.gz43_521145 | AK004557 | Mus musculus adult male lung cDNA, RIKEN full-length enriched library, clone:1200003C23, full insert sequence | 1.20E-05 |
| 1062 | 3665.O19.gz43_521209 | AY036905 | Trichoderma atroviride protein GTPase Tga1 (tga1) gene, complete cds | 2.10E-08 |
| 1063 | 3665.O21.gz43_521241 | U89293 | Homo sapiens MSH4 (HMSH4) mRNA, complete cds | 1.20E-39 |
| 1064 | 3665.O23.gz43_521273 | X00048 | Herpes simplex virus (HSV) type 2 transforming region mtr-2 (map coordinates 0.580 - 0.625) | 6.00E-06 |
| 1065 | 3665.P13.gz43_521114 | Z48796 | H.sapiens Ski-W mRNA for helicase | 1.70E-05 |
| 1066 | 3666.A07.gz43_521387 | AK005546 | Mus musculus adult female placenta cDNA, RIKEN full-length enriched library, clone:1600027G01, full insert sequence | 1.20E-07 |
| 1067 | 3666.A19.gz43_521579 | AB011101 | Homo sapiens mRNA for KIAA0529 protein, partial cds | 5.80E-05 |
| 1068 | 3666.A24.gz43_521659 | AL050208 | Homo sapiens mRNA; cDNA DKFZp586F2323 (from clone DKFZp586F2323) | 2.90E-07 |
| 1069 | 3666.B11.gz43_521452 | X06932 | Petunia hsp70 gene | 3.00E-06 |
| 1070 | 3666.C18.gz43_521565 | Z22672 | H.sapiens cacnl1a3 gene encoding skeletal muscle dhp-receptor alpha 1 subunit | 2.80E-07 |
| 1071 | 3666.D02.gz43_521310 | AJ297538 | Homo sapiens partial RARA gene, intron 2 | 4.00E-06 |
| 1072 | 3666.D11.gz43_521454 | AF057695 | Haemophilus ducreyi strain 35000 putative phosphomannomutase (pmm) gene, partial cds; large supernatant protein 1 (lspA1) gene, complete cds; and putative GMP synthase (guaA) gene, partial cds | 2.43E-04 |
| 1073 | 3666.D15.gz43_521518 | Z66194 | H.sapiens CpG island DNA genomic Mse1 fragment, clone 80b12, forward read cpg80b12.ft1b | 1.70E-66 |
| 1074 | 3666.D16.gz43_521534 | Z66194 | H.sapiens CpG island DNA genomic Mse1 fragment, clone 80b12, forward read cpg80b12.ft1b | 2.10E-37 |
| 1075 | 3666.F22.gz43_521632 | U97062 | Staphylococcus aureus NCTC 8325 SecA (secA) gene, complete cds | 1.20E-08 |
| 1076 | 3666.G12.gz43_521473 | J03901 | Maize pyruvate,orthophosphate dikinase mRNA, complete cds | 1.72E-04 |
| 1077 | 3666.I12.gz43_521475 | AJ225102 | Pinus lambertiana chloroplast DNA containing a SSR Black Hills (Oregon) | 6.40E-10 |
| 1078 | 3666.L01.gz43_521302 | M86227 | Staphylococcus aureus DNA gyrase B subunit (gyrB) RecF homologue (recF) and DNA gyrase A subunit (gyrA) gene, complete cds | 5.00E-06 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 1079 | 3666.L06.gz43_521382 | AF224725 | Trichosurus vulpecula retrovirus TvERV (type D) gag polyprotein (gag), protease (pro), and pol polyprotein (pol) genes, complete cds | 3.30E-08 |
| 1080 | 3666.L11.gz43_521462 | AF147081 | Homo sapiens gamma-glutamyl hydrolase gene, exons 1 and 2 | 3.30E-05 |
| 1081 | 3666.L23.gz43_521654 | AK020701 | Mus musculus 6 days neonate skin cDNA, RIKEN full-length enriched library, clone:A030009B12, full insert sequence | 2.20E-07 |
| 1082 | 3666.M16.gz43_521543 | AF158179 | Drosophila melanogaster strain Canton-S Chiffon-2 (chiffon) mRNA, alternative splice form 2, complete cds | 4.40E-07 |
| 1083 | 3666.N06.gz43_521384 | Z48796 | H.sapiens Ski-W mRNA for helicase | 1.70E-05 |
| 1084 | 3667.A15.gz43_524557 | AF005903 | Monodelphis domestica GTP-binding protein homolog mRNA, partial cds | 7.80E-08 |
| 1085 | 3754.A08.gz43_532949 | AF091502 | Lactobacillus reuteri autoaggregation-mediating protein (aggH) gene, complete cds | 1.00E-06 |
| 1086 | 3754.A13.gz43_533029 | U02695 | Protomelas similis clone PsiI 32 SATA satellite DNA sequence | 7.60E-07 |
| 1087 | 3754.A16.gz43_533077 | AE006577 | Streptococcus pyogenes M1 GAS strain SF370, section 106 of 167 of the complete genome | 9.00E-06 |
| 1088 | 3754.B04.gz43_532886 | S83995 | Pst1 fragment [Chlamydia pneumoniae, Genomic, 474 nt] | 2.00E-06 |
| 1089 | 3754.B05.gz43_532902 | AY008833 | Staphylococcus aureus tcaR-tcaA-tcaB operon, complete sequences | 5.00E-06 |
| 1090 | 3754.B07.gz43_532934 | AF270216 | Staphylococcus epidermidis strain SR1 clone step.1054h11 genomic sequence | 9.50E-07 |
| 1091 | 3754.B08.gz43_532950 | AK007308 | Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:1700128E15, full insert sequence | 7.00E-06 |
| 1092 | 3754.B10.gz43_532982 | AE002807 | Drosophila melanogaster genomic scaffold 142000013385251, complete sequence | 5.40E-05 |
| 1093 | 3754.C22.gz43_533175 | D30612 | Homo sapiens mRNA for repressor protein, partial cds | 4.00E-06 |
| 1094 | 3754.D19.gz43_533128 | L12043 | Plasmodium falciparum unidentified mRNA sequence | 3.00E-06 |
| 1095 | 3754.E12.gz43_533017 | AB062933 | Macaca fascicularis brain cDNA clone:QccE-22249, full insert sequence | 3.60E-07 |
| 1096 | 3754.E20.gz43_533145 | AL138746 | Human DNA sequence from clone RP3-389B13 on chromosome Xq26.2-27.1, complete sequence [Homo sapiens] | 8.30E-10 |
| 1097 | 3754.F01.gz43_532842 | AF086820 | Drosophila melanogaster paired-like homeodomain protein UNC-4 (unc-4) mRNA, complete cds | 8.00E-06 |
| 1098 | 3754.F08.gz43_532954 | S66402 | vascular AT1a angiotensin receptor {exon 1, promoter} [rats, Sprague-Dawley, Genomic, 3477 nt] | 3.10E-05 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 1099 | 3754.F11.gz43_533002 | X57377 | Mouse dilute myosin heavy chain gene for novel heavy chain with unique C-terminal region | 2.10E-05 |
| 1100 | 3754.F15.gz43_533066 | AJ245620 | Homo sapiens CTL1 gene | 2.50E-12 |
| 1101 | 3754.F20.gz43_533146 | AE002426 | Neisseria meningitidis serogroup B strain MC58 section 68 of 206 of the complete genome | 3.70E-05 |
| 1102 | 3754.G03.gz43_532875 | AF002166 | Xenopus laevis Ig mu heavy chain switch region sequence | 1.20E-07 |
| 1103 | 3754.G08.gz43_532955 | X71020 | N.tabacum Npg1 gene for polygalacturonase | 6.80E-07 |
| 1104 | 3754.G18.gz43_533115 | AF126531 | Homo sapiens putative DNA-directed RNA polymerase III C11 subunit gene, complete cds | 1.10E-13 |
| 1105 | 3754.H08.gz43_532956 | L20127 | Rochalimaea henselae antigen (htrA) gene, complete cds | 4.60E-07 |
| 1106 | 3754.I01.gz43_532845 | AK022138 | Homo sapiens cDNA FLJ12076 fis, clone HEMBB1002442, weakly similar to LIN-10 PROTEIN | 3.90E-14 |
| 1107 | 3754.I03.gz43_532877 | AF016653 | Caenorhabditis elegans cosmid C41D7, complete sequence | 2.00E-06 |
| 1108 | 3754.J01.gz43_532846 | U97408 | Caenorhabditis elegans cosmid F48A9 | 4.00E-06 |
| 1109 | 3754.J05.gz43_532910 | Z35484 | Thermoanaerobacter sp. ATCC53627 cgtA gene | 4.00E-06 |
| 1110 | 3754.J10.gz43_532990 | D17094 | Human HepG2 partial cDNA, clone hmd5h04m5 | 5.10E-11 |
| 1111 | 3754.J12.gz43_533022 | Z56695 | H.sapiens CpG island DNA genomic Mse1 fragment, clone 136d4, reverse read cpg136d4.rt1a | 1.00E-06 |
| 1112 | 3754.J24.gz43_533214 | Y12855 | Homo sapiens P2X7 gene, exon 12 and 13 | 2.50E-05 |
| 1113 | 3754.K14.gz43_533055 | L79913 | Xenopus laevis rds/peripherin (rds35) mRNA, complete cds | 5.00E-06 |
| 1114 | 3754.K17.gz43_533103 | AE006251 | Lactococcus lactis subsp. lactis IL1403 section 13 of 218 of the complete genome | 9.00E-06 |
| 1115 | 3754.K20.gz43_533151 | AB047880 | Macaca fascicularis brain cDNA, clone:QnpA-14303 | 1.00E-06 |
| 1116 | 3754.M08.gz43_532961 | X58467 | Human CYP2D7AP pseudogene for cytochrome P450 2D6 | 4.30E-11 |
| 1117 | 3754.N16.gz43_533090 | U33116 | Saccharomyces cerevisiae high copy DNA polymerase suppressor alpha mutation gene (PSP2), complete cds | 1.80E-07 |
| 1118 | 3754.N19.gz43_533138 | AK025312 | Homo sapiens cDNA: FLJ21659 fis, clone COL08743 | 1.40E-07 |
| 1119 | 3754.N22.gz43_533186 | AF081828 | Ixodes hexagonus mitochondrial DNA, complete genome | 4.00E-06 |
| 1120 | 3754.O18.gz43_533123 | Z73229 | S.cerevisiae chromosome XII reading frame ORF YLR057w | 3.00E-06 |
| 1121 | 3754.O23.gz43_533203 | AE006900 | Sulfolobus solfataricus section 259 of 272 of the complete genome | 1.10E-05 |
| 1122 | 3754.P13.gz43_533044 | AF220217 | Homo sapiens rsec15-like protein mRNA, partial cds | 1.80E-10 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 1123 | 3754.P17.gz43_533108 | AJ250862 | Bacillus sp. HIL-Y85/54728 mersacidin biosynthesis gene cluster (mrsK2, mrsR2, mrsF, mrsG, mrsE, mrsA, mrsR1, mrsD, mrsM and mrsT genes) | 1.20E-05 |
| 1124 | 3756.A02.gz43_533237 | AF285594 | Homo sapiens testis protein TEX11 (TEX11) mRNA, complete cds | 1.10E-05 |
| 1125 | 3756.A11.gz43_533381 | U43148 | Human patched homolog (PTC) mRNA, complete cds | 4.00E-06 |
| 1126 | 3756.A13.gz43_533413 | U56861 | Nicotiana plumbaginifolia intergenic region between lhcb1*1 and lhcb1*2 genes | 1.00E-06 |
| 1127 | 3756.B03.gz43_533254 | AF101735 | Pan troglodytes isolate PTOR3A5P olfactory receptor pseudogene, complete sequence | 5.70E-08 |
| 1128 | 3756.B04.gz43_533270 | Z82038 | C.thermosaccharolyticum etfB, etfA, hbd, thlA and actA genes | 1.00E-06 |
| 1129 | 3756.B15.gz43_533446 | M96151 | Mus musculus apolipoprotein B gene sequence | 1.13E-04 |
| 1130 | 3756.B21.gz43_533542 | Z92793 | Caenorhabditis elegans cosmid H15M21, complete sequence | 1.30E-05 |
| 1131 | 3756.B22.gz43_533558 | U43542 | Nicotiana tabacum diphenol oxidase mRNA, complete cds | 2.00E-06 |
| 1132 | 3756.C06.gz43_533303 | AB022085 | Mus musculus Cctz-2 gene for chaperonin containing TCP-1 zeta-2 subunit, exon 5, 6, 7, 8, 9, 10 | 7.00E-05 |
| 1133 | 3756.C16.gz43_533463 | AF143236 | Homo sapiens apoptosis related protein APR 2 mRNA, complete cds | 5.00E-06 |
| 1134 | 3756.D08.gz43_533336 | AB049544 | Porcine enterovirus 10 gene for RNA-dependent RNA polymerase, partial cds | 7.20E-07 |
| 1135 | 3756.D18.gz43_533496 | X53658 | E.coli DNA fragment | 7.60E-08 |
| 1136 | 3756.D24.gz43_533592 | X96861 | H.virescens mRNA for pheromone binding protein | 2.40E-07 |
| 1137 | 3756.E01.gz43_533225 | AF202892 | Mus musculus Kif21a (Kif21a) mRNA, complete cds | 4.00E-06 |
| 1138 | 3756.E06.gz43_533305 | AF139374 | Homo sapiens DIR1 protein (DIR1) gene, complete cds | 8.00E-06 |
| 1139 | 3756.E12.gz43_533401 | AF238884 | Botrytis virus F, complete genome | 8.00E-06 |
| 1140 | 3756.E22.gz43_533561 | U78866 | Arabidopsis thaliana putative arginine-aspartate-rich RNA binding protein (gene1500), (gene1000), and (gene400) genes, complete cds | 5.00E-06 |
| 1141 | 3756.F11.gz43_533386 | D50091 | Drosophila ezoana G-3-P dehydrogenase (alphaGpdh) gene, exon1-8, complete cds | 2.00E-06 |
| 1142 | 3756.F16.gz43_533466 | AJ233973 | Gallus gallus microsatellite DNA GCT028 (CA) repeat | 4.20E-07 |
| 1143 | 3756.G07.gz43_533323 | AE000708 | Aquifex aeolicus section 40 of 109 of the complete genome | 6.00E-05 |
| 1144 | 3756.G12.gz43_533403 | M84731 | Pseudomonas sp. 5-substituted hydantoin racemase (hyuE) gene, complete cds | 1.20E-05 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 1145 | 3756.G14.gz43_533435 | AL116458 | Botrytis cinerea strain T4 cDNA library under conditions of nitrogen deprivation | 6.70E-07 |
| 1146 | 3756.I03.gz43_533261 | U67550 | Methanococcus jannaschii section 92 of 150 of the complete genome | 2.30E-05 |
| 1147 | 3756.J05.gz43_533294 | U11292 | Human Ki nuclear autoantigen mRNA, complete cds | 7.70E-07 |
| 1148 | 3756.K03.gz43_533263 | AF073484 | Homo sapiens MHC class I-related protein MR1 precursor (MR1) gene, signal peptide | 8.00E-06 |
| 1149 | 3756.K07.gz43_533327 | M37499 | Human methylmalonyl CoA mutase (MUT) gene, exon 2 | 2.00E-06 |
| 1150 | 3756.K15.gz43_533455 | AF248820 | Maoricicada campbelli isolate TB-MC-016 tRNA-Asp gene, complete sequence; ATPase subunit 8 gene, complete cds; and ATPase subunit 6 gene, partial cds; mitochondrial genes for mitochondrial products | 7.30E-07 |
| 1151 | 3756.K18.gz43_533503 | M36300 | S.cerevisiae glutamine amidotransferase (TRP3) gene, 3' end | 2.30E-05 |
| 1152 | 3756.K20.gz43_533535 | AY022480 | Oryza sativa microsatellite MRG4805 containing (AGG)X8, genomic sequence | 2.00E-10 |
| 1153 | 3756.L02.gz43_533248 | X03833 | Human gene for interleukin 1 alpha (IL-1 alpha) | 2.80E-12 |
| 1154 | 3756.L03.gz43_533264 | AF244246 | Dysdera sp. MC cytochrome c oxidase I (COI) gene, partial cds; mitochondrial gene for mitochondrial product | 2.70E-07 |
| 1155 | 3756.L19.gz43_533520 | AJ002732 | Schizosaccharomyces pombe mRNA for ribosomal protein l14 | 2.00E-06 |
| 1156 | 3756.M06.gz43_533313 | AK002951 | Mus musculus adult male brain cDNA, RIKEN full-length enriched library, clone:0710001E20, full insert sequence | 3.60E-07 |
| 1157 | 3756.M07.gz43_533329 | AF057708 | Populus balsamifera subsp. trichocarpa PTD protein (PTD) gene, complete cds | 2.60E-07 |
| 1158 | 3756.M20.gz43_533537 | Z35821 | S.cerevisiae chromosome II reading frame ORF YBL060w | 2.00E-06 |
| 1159 | 3756.N18.gz43_533506 | AL591667 | Human DNA sequence from clone RP11-389N9 on chromosome 6, complete sequence [Homo sapiens] | 6.10E-05 |
| 1160 | 3756.N21.gz43_533554 | AK026258 | Homo sapiens cDNA: FLJ22605 fis, clone HSI04743 | 2.00E-06 |
| 1161 | 3756.O03.gz43_533267 | U61347 | Leiophyllum buxifolium ribosomal maturase (matK) gene, chloroplast gene encoding chloroplast protein, complete cds | 4.20E-07 |
| 1162 | 3756.O07.gz43_533331 | AF177871 | Drosophila melanogaster small GTPase RHO1 (Rho1) gene, alternatively spliced products and complete cds | 5.70E-07 |
| 1163 | 3756.O08.gz43_533347 | M60705 | Homo sapiens type I DNA topoisomerase gene, exons 19 and 20 | 6.00E-06 |
| 1164 | 3756.P08.gz43_533348 | M60705 | Homo sapiens type I DNA topoisomerase gene, exons 19 and 20 | 1.00E-05 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 1165 | 3759.C01.gz43_533607 | X71874 | H.sapiens genes for proteasome-like subunit (MECL-1), chymotrypsin-like protease (CTRL-1) and protein serine kinase (PSK-H1) last exon | 4.00E-06 |
| 1166 | 3759.D15.gz43_533832 | AL356790 | Human DNA sequence from clone RP11-238J15 on chromosome 20 Contains ESTs and GSSs. Contains part of the TOM gene for a putative mitochondrial outer membrane protein import receptor similar to yeast pre-mRNA splicing factors Prp1/Zer1 and Prp6, complete> | 1.10E-07 |
| 1167 | 3759.H08.gz43_533724 | M31684 | D.melanogaster cytoskeleton-like bicaudalD protein (BicD) mRNA, complete cds | 2.00E-06 |
| 1168 | 3759.H15.gz43_533836 | AB046001 | Macaca fascicularis brain cDNA, clone:QccE-12738 | 2.60E-07 |
| 1169 | 3759.H17.gz43_533868 | AE000706 | Aquifex aeolicus section 38 of 109 of the complete genome | 1.30E-05 |
| 1170 | 3759.H23.gz43_533964 | AK027088 | Homo sapiens cDNA: FLJ23435 fis, clone HRC12631 | 6.20E-34 |
| 1171 | 3759.I05.gz43_533677 | AF056433 | Homo sapiens clone FBD3 Cri-du-chat critical region mRNA | 1.70E-07 |
| 1172 | 3759.I19.gz43_533901 | Z69666 | Human DNA sequence from cosmid 24F8 from a contig from the tip of the short arm of chromosome 16, spanning 2Mb of 16p13.3. Contains ESTs, repeat polymorphism and CpG island | 2.06E-04 |
| 1173 | 3759.K05.gz43_533679 | L01432 | Soybean calmodulin (SCaM-3) mRNA, complete cds | 4.10E-08 |
| 1174 | 3759.K17.gz43_533871 | Z33340 | M.capricolum DNA for CONTIG MC456 | 4.00E-06 |
| 1175 | 3759.L02.gz43_533632 | U26736 | Caenorhabditis elegans stomatin-like protein MEC-2 (mec-2) gene, complete cds | 3.70E-05 |
| 1176 | 3759.L09.gz43_533744 | M11180 | Transposon Tn917 (complete), macrolide-lincosamide-streptogramin-B (MLS) resistance, complete cds | 1.50E-07 |
| 1177 | 3759.L10.gz43_533760 | AF117022 | Solaria atropurpurea trnL gene, partial sequence; chloroplast gene for chloroplast product | 4.40E-07 |
| 1178 | 3759.L15.gz43_533840 | U22657 | Mus musculus genomic locus related to cellular morphology | 1.60E-05 |
| 1179 | 3759.L24.gz43_533984 | AK022990 | Homo sapiens cDNA FLJ12928 fis, clone NT2RP2004767 | 7.60E-10 |
| 1180 | 3759.M19.gz43_533905 | M96324 | Lycopersicon esculentum Ca2+-ATPase gene, complete cds | 2.50E-05 |
| 1181 | 3759.N08.gz43_533730 | AK005546 | Mus musculus adult female placenta cDNA, RIKEN full-length enriched library, clone:1600027G01, full insert sequence | 1.30E-07 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 1182 | 3759.N16.gz43_533858 | AB014079 | Homo sapiens genomic DNA, chromosome 6p21.3, HLA class I region, Cosmid clone:TY1E11, complete sequence | 3.80E-12 |
| 1183 | 3759.N23.gz43_533970 | AK018377 | Mus musculus 16 days embryo lung cDNA, RIKEN full-length enriched library, clone:8430403M08, full insert sequence | 5.70E-07 |
| 1184 | 3759.O16.gz43_533859 | AE000918 | Methanobacterium thermoautotrophicum from bases 1444576 to 1460617 (section 124 of 148) of the complete genome | 1.40E-05 |
| 1185 | 3759.P03.gz43_533652 | L06066 | Saccharomyces cerevisiae PET117 polypeptide (PET117) gene, complete cds | 5.90E-07 |
| 1186 | 3759.P13.gz43_533812 | X89414 | A.thaliana DNA for pyrroline-5-carboxylase synthetase gene | 5.00E-06 |
| 1187 | 3759.P15.gz43_533844 | X66979 | X.laevis mRNA XLFLI | 1.50E-05 |
| 1188 | 3759.P17.gz43_533876 | AF039313 | Moraxella catarrhalis strain LES-1 transferrin binding protein B (tbpB) gene, complete cds | 2.00E-06 |
| 1189 | 3762.A09.gz43_534117 | AE000496 | Escherichia coli K12 MG1655 section 386 of 400 of the complete genome | 1.63E-04 |
| 1190 | 3762.A16.gz43_534229 | X98371 | D.subobscura sex-lethal gene | 7.00E-06 |
| 1191 | 3762.A19.gz43_534277 | U95019 | Human voltage-dependent calcium channel beta-2c subunit mRNA, complete cds | 6.10E-07 |
| 1192 | 3762.A20.gz43_534293 | M10014 | Homo sapiens map 4q28 fibrinogen (FGG) gene, alternative splice products, complete cds | 8.00E-06 |
| 1193 | 3762.B05.gz43_534054 | J05614 | Human proliferating cell nuclear antigen (PCNA) gene, promoter region | 1.40E-05 |
| 1194 | 3762.B15.gz43_534214 | AJ297559 | Homo sapiens partial PIK3CB gene for phosphatidylinositol 3-kinase catalytic subunit p110beta, exons 15-17 | 2.50E-05 |
| 1195 | 3762.C20.gz43_534295 | M58580 | Rabbit angiotensin-converting enzyme (ACE) gene, 5' end | 3.10E-05 |
| 1196 | 3762.C23.gz43_534343 | L27146 | Human neurofibromatosis 2 (NF2) gene, exon 16 | 1.00E-06 |
| 1197 | 3762.D03.gz43_534024 | U51305 | Triticum aestivum alpha-gliadin storage protein pseudogene, complete cds | 1.40E-05 |
| 1198 | 3762.D04.gz43_534040 | AF263274 | Chionodraco rastrospinosus isolate Cra7 alpha tubulin mRNA, complete cds | 3.50E-07 |
| 1199 | 3762.D18.gz43_534264 | M94764 | Glycine max cv. Dare nodulin 26 gene fragment | 2.50E-05 |
| 1200 | 3762.D19.gz43_534280 | AE001446 | Helicobacter pylori, strain J99 section 7 of 132 of the complete genome | 3.30E-05 |
| 1201 | 3762.D22.gz43_534328 | M73962 | Bovine pregnancy-associated glycoprotein 1 mRNA, complete cds | 4.00E-06 |
| 1202 | 3762.E01.gz43_533993 | X63746 | S.cerevisiae rpc34 and fun34 genes for DNA dependant RNA polymerase c (III) | 4.00E-06 |
| 1203 | 3762.E10.gz43_534137 | Z74847 | S.cerevisiae chromosome XV reading frame ORF YOL105c | 1.00E-05 |

Table 8

| SEQ ID | SEQ NAME | ACCESSION | GENBANK DESCRIPTION | GENBANK SCORE |
|---|---|---|---|---|
| 1204 | 3762.E15.gz43_534217 | AF207841 | Pyricularia grisea AVR-Pita (AVR-Pita) gene, complete cds | 2.20E-09 |
| 1205 | 3762.E23.gz43_534345 | M58600 | Human heparin cofactor II (HCF2) gene, exons 1 through 5 | 3.60E-37 |
| 1206 | 3762.F08.gz43_534106 | Z47066 | Human cosmid Qc14G3 from Xq28 contains STSs | 3.10E-09 |
| 1207 | 3762.F22.gz43_534330 | AY034974 | Arabidopsis thaliana unknown protein (F24J8.3) mRNA, complete cds | 4.20E-07 |
| 1208 | 3762.G18.gz43_534267 | Z28150 | S.cerevisiae chromosome XI reading frame ORF YKL150w | 2.00E-06 |
| 1209 | 3762.H12.gz43_534172 | AF370230 | Arabidopsis thaliana unknown protein (T21P5_16/AT3g03420) mRNA, complete cds | 6.60E-08 |
| 1210 | 3762.I07.gz43_534093 | U19569 | Human squamous cell carcinoma antigen (SCCA2) gene, exon 1 | 4.60E-07 |
| 1211 | 3762.J03.gz43_534030 | U22421 | Mus musculus obesity protein (ob) gene, complete cds | 5.30E-07 |
| 1212 | 3762.J18.gz43_534270 | AB027966 | Schizosaccharomyces pombe gene for Hypothetical protein, partial cds, clone:TB89 | 2.30E-08 |
| 1213 | 3762.K02.gz43_534015 | AF273762 | Homo sapiens 3-hydroxy-3-methylglutaryl-coenzyme reductase gene, exon 15 | 4.40E-14 |
| 1214 | 3762.K20.gz43_534303 | K01464 | Rat cardiac alpha-myosin heavy chain gene, 5' flank, 1st 3 exons | 3.00E-06 |
| 1215 | 3762.L18.gz43_534272 | Z49438 | S.cerevisiae chromosome X reading frame ORF YJL163c | 4.00E-06 |
| 1216 | 3762.L20.gz43_534304 | XM_030040 | Homo sapiens similar to KIAA0877 protein (H. sapiens) (LOC90219), mRNA | 3.00E-06 |
| 1217 | 3762.M04.gz43_534049 | AF002237 | Anopheles gambiae clone 227 mRNA sequence | 4.00E-06 |
| 1218 | 3762.M17.gz43_534257 | M29688 | S.cerevisiae PMS1 gene encoding DNA mismatch repair protein, complete cds | 1.40E-08 |
| 1219 | 3762.M23.gz43_534353 | M20006 | Chicken tumor 10 c-myc DNA, exons 2 and 3 | 2.90E-09 |
| 1220 | Clu1014734.con_1 | AB027966 | Schizosaccharomyces pombe gene for Hypothetical protein, partial cds, clone:TB89 | 3.00E-08 |
| 1221 | Clu1036845.con_1 | M34429 | Human PVT-IGLC fusion protein mRNA, 5' end | 1.37E-03 |

**Table 9**

| SEQ ID | SEQ NAME | PFAM NAME | PFAM DESCRIPTION | SCORE | START | END |
|---|---|---|---|---|---|---|
| **89** | 3547.D19.GZ43_505986 | DC1 | DC1 domain | 30.64 | 411 | 493 |
| **137** | 3550.G02.GZ43_506101 | rvt | Reverse transcriptase (RNA-dependent DNA polymerase) | 47.32 | 321 | 611 |
| **321** | 3562.B22.GZ43_507952 | 7tm_1 | 7 transmembrane receptor (rhodopsin family) | 37.16 | 154 | 479 |
| **321** | 3562.B22.GZ43_507952 | Bowman-Birk_leg | Bowman-Birk serine protease inhibitor family | 45.92 | 292 | 450 |
| **321** | 3562.B22.GZ43_507952 | Cation_efflux | Cation efflux family | 33.32 | 225 | 380 |
| **358** | 3565.E16.GZ43_508243 | AP_endonucleas1 | AP endonuclease family 1 | 38.16 | 406 | 577 |
| **413** | 3571.A08.GZ43_508897 | oxidored_q1 | NADH-Ubiquinone/plastoquinone (complex I), various chains | 30.04 | 297 | 393 |
| **417** | 3571.B13.GZ43_508978 | EGF | EGF-like domain | 38.88 | 243 | 355 |
| **418** | 3571.B22.GZ43_509122 | EGF | EGF-like domain | 38.88 | 243 | 355 |
| **431** | 3571.H10.GZ43_508936 | WW | WW domain | 54.92 | 487 | 576 |
| **591** | 3583.H13.GZ43_510520 | Sre | C. elegans Sre G protein-coupled chemoreceptor | 30.36 | 282 | 485 |
| **638** | 3590.J21.GZ43_512427 | bZIP | bZIP transcription factor | 33.68 | 166 | 308 |
| **645** | 3590.M03.GZ43_512142 | protamine_P1 | Protamine P1 | 35.88 | 268 | 437 |
| **774** | 3608.L14.gz43_514237 | Transposase_22 | L1 transposable element | 62.12 | 491 | 616 |
| **836** | 3617.P12.gz43_515361 | AP_endonucleas1 | AP endonuclease family 1 | 39.84 | 63 | 254 |
| **905** | 3632.O06.gz43_517407 | 60s_ribosomal | 60s Acidic ribosomal protein | 38.04 | 276 | 444 |
| **905** | 3632.O06.gz43_517407 | 60s_ribosomal | 60s Acidic ribosomal protein | 36.44 | 13 | 98 |
| **995** | 3662.H23.gz43_520114 | Glycoprotein_G | Pneumovirus attachment glycoprotein G | 43.04 | 21 | 297 |
| **995** | 3662.H23.gz43_520114 | Metallothio_5 | Metallothionein family 5 | 47.88 | 231 | 345 |
| **995** | 3662.H23.gz43_520114 | squash | Squash family serine protease inhibitor | 34.6 | 222 | 301 |
| **995** | 3662.H23.gz43_520114 | Syndecan | Syndecan domain | 35.36 | 1 | 308 |
| **1012** | 3663.G01.gz43_520145 | KRAB | KRAB box | 95.08 | 424 | 484 |
| **1217** | 3762.M04.gz43_534049 | protamine_P1 | Protamine P1 | 33.16 | 293 | 468 |

Table 10

| SEQ ID | SEQ NAME | PFAM NAME | PFAM DESCRIPTION | SCORE | START | END |
|---|---|---|---|---|---|---|
| 1486 | NTP_004511S11.3_4 | Armadillo_seg | Armadillo/beta-catenin-like repeat | 1.8E-95 | 142 | 184 |
| 1486 | NTP_004511S11.3_4 | Armadillo_seg | Armadillo/beta-catenin-like repeat | 1.8E-95 | 186 | 226 |
| 1486 | NTP_004511S11.3_4 | Armadillo_seg | Armadillo/beta-catenin-like repeat | 1.8E-95 | 228 | 269 |
| 1486 | NTP_004511S11.3_4 | Armadillo_seg | Armadillo/beta-catenin-like repeat | 1.8E-95 | 271 | 311 |
| 1486 | NTP_004511S11.3_4 | Armadillo_seg | Armadillo/beta-catenin-like repeat | 1.8E-95 | 313 | 353 |
| 1486 | NTP_004511S11.3_4 | Armadillo_seg | Armadillo/beta-catenin-like repeat | 1.8E-95 | 355 | 395 |
| 1486 | NTP_004511S11.3_4 | Armadillo_seg | Armadillo/beta-catenin-like repeat | 1.8E-95 | 397 | 437 |
| 1486 | NTP_004511S11.3_4 | Armadillo_seg | Armadillo/beta-catenin-like repeat | 1.8E-95 | 440 | 480 |
| 1486 | NTP_004511S11.3_4 | Armadillo_seg | Armadillo/beta-catenin-like repeat | 1.8E-95 | 142 | 184 |
| 1486 | NTP_004511S11.3_4 | Armadillo_seg | Armadillo/beta-catenin-like repeat | 1.8E-95 | 186 | 226 |
| 1486 | NTP_004511S11.3_4 | Armadillo_seg | Armadillo/beta-catenin-like repeat | 1.8E-95 | 228 | 269 |
| 1486 | NTP_004511S11.3_4 | Armadillo_seg | Armadillo/beta-catenin-like repeat | 1.8E-95 | 271 | 311 |
| 1486 | NTP_004511S11.3_4 | Armadillo_seg | Armadillo/beta-catenin-like repeat | 1.8E-95 | 313 | 353 |
| 1486 | NTP_004511S11.3_4 | Armadillo_seg | Armadillo/beta-catenin-like repeat | 1.8E-95 | 355 | 395 |
| 1486 | NTP_004511S11.3_4 | Armadillo_seg | Armadillo/beta-catenin-like repeat | 1.8E-95 | 397 | 437 |
| 1486 | NTP_004511S11.3_4 | Armadillo_seg | Armadillo/beta-catenin-like repeat | 1.8E-95 | 440 | 480 |
| 1486 | NTP_004511S11.3_4 | IBB | Importin beta binding domain | 5.8E-37 | 35 | 124 |
| 1486 | NTP_004511S11.3_4 | IBB | Importin beta binding domain | 5.8E-37 | 35 | 124 |
| 1497 | NTP_007592S2.3_10 | histone | Core histone H2A/H2B/H3/H4 | 1.2E-10 | 2 | 97 |
| 1497 | NTP_007592S2.3_10 | histone | Core histone H2A/H2B/H3/H4 | 1.2E-10 | 2 | 97 |
| 1500 | NTP_007867S7.3_3 | GTF2I | GTF2I-like repeat | 7.2E-76 | 106 | 171 |
| 1500 | NTP_007867S7.3_3 | GTF2I | GTF2I-like repeat | 7.2E-76 | 295 | 370 |
| 1500 | NTP_007867S7.3_3 | GTF2I | GTF2I-like repeat | 7.2E-76 | 106 | 171 |
| 1500 | NTP_007867S7.3_3 | GTF2I | GTF2I-like repeat | 7.2E-76 | 295 | 370 |
| 1501 | NTP_007867S8.3_1 | GTF2I | GTF2I-like repeat | 7.2E-76 | 122 | 187 |
| 1501 | NTP_007867S8.3_1 | GTF2I | GTF2I-like repeat | 7.2E-76 | 311 | 386 |
| 1501 | NTP_007867S8.3_1 | GTF2I | GTF2I-like repeat | 7.2E-76 | 122 | 187 |
| 1501 | NTP_007867S8.3_1 | GTF2I | GTF2I-like repeat | 7.2E-76 | 311 | 386 |
| 1507 | NTP_008858S2.3_2 | GST_N | Glutathione S-transferase, N-terminal domain | 4.6E-11 | 21 | 95 |
| 1507 | NTP_008858S2.3_2 | GST_N | Glutathione S-transferase, N-terminal domain | 4.6E-11 | 21 | 95 |
| 1510 | NTP_009526S2.3_3 | CBS | CBS domain | 4.8E-43 | 30 | 84 |

Table 10

| SEQ ID | SEQ NAME | PFAM NAME | PFAM DESCRIPTION | SCORE | START | END |
|---|---|---|---|---|---|---|
| 1510 | NTP_009526S2.3_3 | CBS | CBS domain | 4.8E-43 | 111 | 165 |
| 1510 | NTP_009526S2.3_3 | CBS | CBS domain | 4.8E-43 | 186 | 239 |
| 1510 | NTP_009526S2.3_3 | CBS | CBS domain | 4.8E-43 | 258 | 311 |
| 1510 | NTP_009526S2.3_3 | CBS | CBS domain | 4.8E-43 | 30 | 84 |
| 1510 | NTP_009526S2.3_3 | CBS | CBS domain | 4.8E-43 | 111 | 165 |
| 1510 | NTP_009526S2.3_3 | CBS | CBS domain | 4.8E-43 | 186 | 239 |
| 1510 | NTP_009526S2.3_3 | CBS | CBS domain | 4.8E-43 | 258 | 311 |
| 1511 | NTP_009526S2.3_5 | CBS | CBS domain | 4.8E-43 | 30 | 84 |
| 1511 | NTP_009526S2.3_5 | CBS | CBS domain | 4.8E-43 | 111 | 165 |
| 1511 | NTP_009526S2.3_5 | CBS | CBS domain | 4.8E-43 | 186 | 239 |
| 1511 | NTP_009526S2.3_5 | CBS | CBS domain | 4.8E-43 | 258 | 311 |
| 1511 | NTP_009526S2.3_5 | CBS | CBS domain | 4.8E-43 | 30 | 84 |
| 1511 | NTP_009526S2.3_5 | CBS | CBS domain | 4.8E-43 | 111 | 165 |
| 1511 | NTP_009526S2.3_5 | CBS | CBS domain | 4.8E-43 | 186 | 239 |
| 1511 | NTP_009526S2.3_5 | CBS | CBS domain | 4.8E-43 | 258 | 311 |
| 1514 | NTP_010018S2.3_5 | DAG_PE-bind | Phorbol esters/diacylglycerol binding domain (C1 domain) | 7.7E-23 | 154 | 203 |
| 1514 | NTP_010018S2.3_5 | DAG_PE-bind | Phorbol esters/diacylglycerol binding domain (C1 domain) | 7.7E-23 | 387 | 426 |
| 1514 | NTP_010018S2.3_5 | DAG_PE-bind | Phorbol esters/diacylglycerol binding domain (C1 domain) | 7.7E-23 | 154 | 203 |
| 1514 | NTP_010018S2.3_5 | DAG_PE-bind | Phorbol esters/diacylglycerol binding domain (C1 domain) | 7.7E-23 | 387 | 426 |
| 1518 | NTP_010757S4.3_2 | T-box | T-box | 5.5E-114 | 935 | 1099 |
| 1518 | NTP_010757S4.3_2 | T-box | T-box | 5.5E-114 | 1142 | 1160 |
| 1518 | NTP_010757S4.3_2 | T-box | T-box | 5.5E-114 | 935 | 1099 |
| 1518 | NTP_010757S4.3_2 | T-box | T-box | 5.5E-114 | 1142 | 1160 |
| 1520 | NTP_011130S2.3_3 | GATA | GATA zinc finger | 1.5E-11 | 159 | 198 |
| 1520 | NTP_011130S2.3_3 | GATA | GATA zinc finger | 1.5E-11 | 159 | 198 |
| 1523 | NTP_011430S6.3_6 | cadherin | Cadherin domain | 7.4E-61 | 174 | 270 |
| 1523 | NTP_011430S6.3_6 | cadherin | Cadherin domain | 7.4E-61 | 284 | 390 |

Table 10

| SEQ ID | SEQ NAME | PFAM NAME | PFAM DESCRIPTION | SCORE | START | END |
|---|---|---|---|---|---|---|
| 1523 | NTP_011430S6.3_6 | cadherin | Cadherin domain | 7.4E-61 | 405 | 495 |
| 1523 | NTP_011430S6.3_6 | cadherin | Cadherin domain | 7.4E-61 | 174 | 270 |
| 1523 | NTP_011430S6.3_6 | cadherin | Cadherin domain | 7.4E-61 | 284 | 390 |
| 1523 | NTP_011430S6.3_6 | cadherin | Cadherin domain | 7.4E-61 | 405 | 495 |
| 1525 | NTP_017582S2.3_6 | HMG_box | HMG (high mobility group) box | 6.8E-09 | 34 | 92 |
| 1525 | NTP_017582S2.3_6 | HMG_box | HMG (high mobility group) box | 6.8E-09 | 34 | 92 |
| 1542 | NTP_026331S1.1_1 | GTF2I | GTF2I-like repeat | 7.2E-76 | 106 | 171 |
| 1542 | NTP_026331S1.1_1 | GTF2I | GTF2I-like repeat | 7.2E-76 | 295 | 370 |
| 1542 | NTP_026331S1.1_1 | GTF2I | GTF2I-like repeat | 7.2E-76 | 106 | 171 |
| 1542 | NTP_026331S1.1_1 | GTF2I | GTF2I-like repeat | 7.2E-76 | 295 | 370 |

Table 11

| Pt ID | Path ID | Grp | Anatom Loc | Size | Grade | Histo Grade | Local Invasion | Lymph Met | Lymph Met Incid | Reg Lymph Grade | Dist Met & Loc | Dist Met Grade | Comment |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 15 | 21 | III | Ascending colon | 4.0 | T3 | G2 | Extending into subserosal adipose tissue | Pos | 3/8 | N1 | Neg | MX | invasive adenocarcinoma, moderately differentiated; focal perineural invasion is seen |
| 52 | 71 | II | Cecum | 9.0 | T3 | G3 | Invasion through muscularis propria, subserosal involvement; ileocec. valve involvement | Neg | 0/12 | N0 | Neg | M0 | Hyperplastic polyp in appendix. |
| 121 | 140 | II | Sigmoid | 6 | T4 | G2 | Invasion of muscularis propria into serosa, involving submucosa of urinary bladder | Neg | 0/34 | N0 | Neg | M0 | Perineural invasion; donut anastomosis Neg. One tubulovillous and one tubular adenoma with no high grade dysplasia. |
| 125 | 144 | II | Cecum | 6 | T3 | G2 | Invasion through the muscularis propria into suserosal adipose tissue. Ileocecal junction. | Neg | 0/19 | N0 | Neg | M0 | patient history of metastatic melanoma |
| 128 | 147 | III | Transverse colon | 5.0 | T3 | G2 | Invasion of muscularis propria into percolonic fat | Pos | 1/5 | N1 | Neg | M0 | |

Table 11

| Pt ID | Path ID | Grp | Anatom Loc | Size | Grade | Histo Grade | Local Invasion | Lymph Met | Lymph Met Incid | Reg Lymph Grade | Dist Met & Loc | Dist Met Grade | Comment |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 130 | 149 | | Splenic flexure | 5.5 | T3 | | through wall and into surrounding adipose tissue | Pos | 10/24 | N2 | Neg | M1 | |
| 133 | 152 | II | Rectum | 5.0 | T3 | G2 | Invasion through muscularis propria into non-peritonealized pericolic tissue; gross configuration is annular. | Neg | 0/9 | N0 | Neg | M0 | Small separate tubular adenoma (0.4 cm) |
| 141 | 160 | IV | Cecum | 5.5 | T3 | G2 | Invasion of muscularis propria into pericolonic adipose tissue, but not through serosa. Arising from tubular adenoma. | Pos | 7/21 | N2 | Pos - Liver | M1 | Perineural invasion identified adjacent to metastatic adenocarcinom a. |
| 156 | 175 | III | Hepatic flexure | 3.8 | T3 | G2 | Invasion through mucsularis propria into subserosa/pericolic adipose, no serosal involvement. Gross configuration annular. | Pos | 2/13 | N1 | Neg | M0 | Separate tubolovillous and tubular adenomas |

Table 11

| Pt ID | Path ID | Grp | Anatom Loc | Size | Grade | Histo Grade | Local Invasion | Lymph Met | Lymph Met Incid | Reg Lymph Grade | Dist Met & Loc | Dist Met Grade | Comment |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 228 | 247 | III | Rectum | 5.8 | T3 | G2 to G3 | Invasion through muscularis propria to involve subserosal, perirectoal adipose, and serosa | Pos | 1/8 | N1 | Neg | MX | Hyperplastic polyps |
| 264 | 283 | II | Ascending colon | 5.5 | T3 | G2 | Invasion through muscularis propria into subserosal adipose tissue. | Neg | 0/10 | N0 | Neg | M0 | Tubulovillous adenoma with high grade dysplasia |
| 266 | 285 | III | Transverse colon | 9 | T3 | G2 | Invades through muscularis propria to involve pericolonic adipose, extends to serosa. | Neg | 0/15 | N1 | Pos - Mesenteric deposit | MX | |
| 268 | 287 | I | Cecum | 6.5 | T2 | G2 | Invades full thickness of muscularis propria, but mesenteric adipose free of malignancy | Neg | 0/12 | N0 | Neg | M0 | |
| 278 | 297 | III | Rectum | 4 | T3 | G2 | Invasion into perirectal adipose tissue. | Pos | 7/10 | N2 | Neg | M0 | Descending colon polyps, no HGD or carcinoma identified.. |

**Table 11**

| Pt ID | Path ID | Grp | Anatom Loc | Size | Grade | Histo Grade | Local Invasion | Lymph Met | Lymph Met Incid | Reg Lymph Grade | Dist Met & Loc | Dist Met Grade | Comment |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 296 | 315 | III | Cecum | 5.5 | T3 | G2 | Invasion through muscularis propria and invades pericolic adipose tissue. Ileocecal junction. | Pos | 2/12 | N1 | Neg | M0 | Tubulovillous adenoma (2.0 cm) with no high grade dysplasia. Neg. liver biopsy. |
| 339 | 358 | II | Rectosigm oid | 6 | T3 | G2 | Extends into perirectal fat but does not reach serosa | Neg | 0/6 | N0 | Neg | M0 | 1 hyperplastic polyp identified |
| 341 | 360 | II | Ascending colon | 2 cm inva sive | T3 | G2 | Invasion through muscularis propria to involve pericolonic fat. Arising from villous adenoma. | Neg | 0/4 | N0 | Neg | MX | |
| 356 | 375 | II | Sigmoid | 6.5 | T3 | G2 | Through colon wall into subserosal adipose tissue. No serosal spread seen. | Neg | 0/4 | N0 | Neg | M0 | |
| 360 | 412 | III | Ascending colon | 4.3 | T3 | G2 | Invasion thru muscularis propria to pericolonic fat | Pos | 1/5 | N1 | Neg | M0 | Two mucosal polyps |
| 392 | 444 | IV | Ascending colon | 2 | T3 | G2 | Invasion through muscularis propria into subserosal adipose tissue, not serosa. | Pos | 1/6 | N1 | Pos - Liver | M1 | Tumor arising at prior ileocolic surgical anastomosis. |

**Table 11**

| Pt ID | Path ID | Grp | Anatom Loc | Size | Grade | Histo Grade | Local Invasion | Lymph Met | Lymph Met Incid | Reg Lymph Grade | Dist Met & Loc | Dist Met Grade | Comment |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 393 | 445 | II | Cecum | 6.0 | T3 | G2 | Cecum, invades through muscularis propria to involve subserosal adipose tissue but not serosa. | Neg | 0/21 | N0 | Neg | M0 | |
| 413 | 465 | IV | Cecum | 4.8 | T3 | G2 | Invasive through muscularis to involve periserosal fat; abutting ileocecal junction. | Neg | 0/7 | N0 | Pos - Liver | M1 | rediagnosis of oophorectomy path to metastatic colon cancer. |
| 505 | 383 | IV | | 7.5 | T3 | G2 | Invasion through muscularis propria involving pericolic adipose, serosal surface uninvolved | Pos | 2/17 | N1 | Pos - Liver | M1 | Anatomical location of primary not notated in report. Evidence of chronic colitis. |
| 517 | 395 | IV | Sigmoid | 3 | T3 | G2 | penetrates muscularis propria, involves pericolonic fat. | Pos | 6/6 | N2 | Neg | M0 | No mention of distant met in report |

Table 11

| Pt ID | Path ID | Grp | Anatom Loc | Size | Grade | Histo Grade | Local Invasion | Lymph Met | Lymph Met Incid | Reg Lymph Grade | Dist Met & Loc | Dist Met Grade | Comment |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 534 | 553 | II | Ascending colon | 12 | T3 | G3 | Invasion through the muscularis propria involving pericolic fat. Serosa free of tumor. | Neg | 0/8 | N0 | Neg | M0 | Omentum with fibrosis and fat necrosis. Small bowel with acute and chronic serositis, focal abscess and adhesions. |
| 546 | 565 | IV | Ascending colon | 5.5 | T3 | G2 | Invasion through muscularis propria extensively through submucosal and extending to serosa. | Pos | 6/12 | N2 | Pos - Liver | M1 | |
| 577 | 596 | II | Cecum | 11.5 | T3 | G2 | Invasion through the bowel wall, into suberosal adipose. Serosal surface free of tumor. | Neg | 0/58 | N0 | Neg | M0 | Appendix dilated and fibrotic, but not involved by tumor |

Table 11

| Pt ID | Path ID | Grp | Anatom Loc | Size | Grade | Histo Grade | Local Invasion | Lymph Met | Lymph Met Incid | Reg Lymph Grade | Dist Met & Loc | Dist Met Grade | Comment |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 695 | 714 | II | Cecum | 14.0 | T3 | G2 | extending through bowel wall into serosal fat | Neg | 0/22 | N0 | Neg | MX | moderately differentiated adenocarcinoma with mucinous diferentiation (% not stated), tubular adenoma and hyperplstic polyps present, |
| 784 | 803 | IV | Ascending colon | 3.5 | T3 | G3 | through muscularis propria into pericolic soft tissues | Pos | 5/17 | N2 | Pos - Liver | M1 | invasive poorly differentiated adenosquamous carcinoma |
| 786 | 805 | IV | Descending colon | 9.5 | T3 | G2 | through muscularis propria into pericolic fat, but not at serosal surface | Neg | 0/12 | N0 | Pos - Liver | M1 | moderately differentiated invasive adenocarcinoma |
| 787 | 806 | II | Rectosigmoid | 2.5 | T3 | G2-G3 | Invasion of muscularis propria into soft tissue | Neg | | N0 | Neg | MX | Peritumoral lymphocytic response; 5 LN examined in pericolic fat, no metastatases observed. |
| 789 | 808 | IV | Cecum | 5.0 | T3 | G2-G3 | Extending through muscularis propria into pericolonic fat | Pos | 5/10 | N2 | Pos - Liver | M1 | Three fungating lesions examined. |

Table 11

| Pt ID | Path ID | Grp | Anatom Loc | Size | Grade | Histo Grade | Local Invasion | Lymph Met | Lymph Met Incid | Reg Lymph Grade | Dist Met & Loc | Dist Met Grade | Comment |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 790 | 809 | IV | Rectum | 6.8 | T3 | G1-G2 | Invading through muscularis propria into perirectal fat | Pos | 3/13 | N1 | Pos - Liver | M1 | |
| 791 | 810 | IV | Ascending colon | 5.8 | T3 | G3 | Through the muscularis propria into pericolic fat | Pos | 13/25 | N2 | Pos - Liver | M1 | poorly differentiated invasive colonic adenocarcinoma |
| 888 | 908 | IV | Ascending colon | 2.0 | T2 | G1 | Into muscularis propria | Pos | 3/21 | N0 | Pos - Liver | M1 | well to moderately differentiated adenocarcinomas; this patient has tumors of the ascending colon and the sigmoid colon |
| 889 | 909 | IV | Cecum | 4.8 | T3 | G2 | Through muscularis propria int subserosal tissue | Pos | 1/4 | N1 | Pos - Liver | M1 | moderately differentiated adenocarcinoma |
| 890 | 910 | IV | Ascending colon | | T3 | G2 | Through muscularis propria into subserosa. | Pos | 11/15 | N2 | Pos - Liver | M1 | |
| 891 | 911 | IV | Rectum | 5.2 | T3 | G2 | Invasion through muscularis propria into perirectal soft tissue | Pos | 4/15 | N2 | Pos - Liver | M1 | Perineural invasion present. |

**Table 11**

| Pt ID | Path ID | Grp | Anatom Loc | Size | Grade | Histo Grade | Local Invasion | Lymph Met | Lymph Met Incid | Reg Lymph Grade | Dist Met & Loc | Dist Met Grade | Comment |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 892 | 912 | IV | Sigmoid | 5.0 | T3 | G2 | Invasion into pericolic sort tissue. Tumor focally invading skeletal muscle attached to colon. | Pos | 1/28 | N1 | Pos - Liver, left and right lobe, omentum | M1 | Perineural invasion present, extensive. Patient with a history of colon cancer. |
| 893 | 913 | IV | Transverse colon | 6.0 | T3 | G2-G3 | Through muscularis propria into pericolic fat | Pos | 14/17 | N2 | Pos - Liver | M1 | Perineural invasion focally present. Omentum mass, but resection with no tumor identified. |
| 989 | 1009 | IV | Sigmoid | 6.0 | T3 | G2 | Invasion through colon wall and focally involving subserosal tissue. | Pos | 1/7 | N1 | Pos - Liver | M1 | Primary adenocarcinoma arising from tubulovillous adenoma. |

Table 13

| SEQ ID | SEQ NAME | CLONE ID | BREAST PATIENTS >=2x | BREAST NUM RATIOS | COLON PATIENTS >=2x | COLON NUM RATIOS | COLON UM >=2x | COLON UM NUM RATIOS |
|---|---|---|---|---|---|---|---|---|
| 56 | 3544.G06.GZ43_505397 | M00084443A:E10 | | | 50 | 8 | | |
| 287 | 3559.B18.GZ43_507504 | M00084700A:C10 | | | 41.025641 | 39 | 33.33333 | 27 |
| 621 | 3590.D19.GZ43_512389 | M00085031B:E03 | | | 37.5 | 8 | | |
| 683 | 3596.P03.GZ43_512529 | M00085171D:F05 | | | 70 | 40 | 60.71429 | 28 |
| 706 | 3599.K02.GZ43_512892 | M00085222D:D07 | | | 70 | 40 | 60.71429 | 28 |
| 1059 | 3665.O06.gz43_521001 | M00086277B:E06 | | | 57.5 | 40 | 50 | 12 |
| 1150 | 3756.K15.gz43_533455 | M00085835B:E11 | | | 35.2941176 | 34 | 35.71429 | 28 |
| 1156 | 3756.M06.gz43_533313 | M00085815C:E11 | 47.0588235 | 17 | | | | |
| 1187 | 3759.P15.gz43_533844 | M00085100B:C12 | | | 63.4146341 | 41 | | |
| 1416 | NT_007592S2.3_10 | | 50 | 10 | | | | |
| 1427 | NT_009296S1.3_1 | | | | | | 42.9 | 28 |
| 1427 | NT_009296S1.3_1 | | | | 46.2 | 39 | 33.3 | 27 |
| 1427 | NT_009296S1.3_1 | | | | 48.7 | 39 | 44.4 | 27 |
| 1444 | NT_017582S2.3_6 | | | | 61.5 | 39 | 55.6 | 27 |
| 1444 | NT_017582S2.3_6 | | | | 61.5 | 39 | 55.6 | 27 |

**Table 15**

| Library ID | CMCC Number | CloneId | NRRL Number |
|---|---|---|---|
| ES219 | 5471 | M00084879B:E01 | B-30523 |
| ES219 | 5471 | M00083819B:E10 | B-30523 |
| ES219 | 5471 | M00084942C:B10 | B-30523 |
| ES219 | 5471 | M00084704C:B09 | B-30523 |
| ES219 | 5471 | M00084887C:C07 | B-30523 |
| ES219 | 5471 | M00084976B:A08 | B-30523 |
| ES219 | 5471 | M00085011B:A01 | B-30523 |
| ES219 | 5471 | M00084961A:C07 | B-30523 |
| ES219 | 5471 | M00084960D:D02 | B-30523 |
| ES219 | 5471 | M00084973A:B06 | B-30523 |
| ES219 | 5471 | M00084928D:F06 | B-30523 |
| ES219 | 5471 | M00084968C:D10 | B-30523 |
| ES219 | 5471 | M00084973A:B06 | B-30523 |
| ES219 | 5471 | M00084966A:A08 | B-30523 |
| ES219 | 5471 | M00084919C:B04 | B-30523 |
| ES219 | 5471 | M00085003C:D03 | B-30523 |
| ES219 | 5471 | M00084968A:D01 | B-30523 |
| ES219 | 5471 | M00084969D:C11 | B-30523 |
| ES219 | 5471 | M00084899D:B01 | B-30523 |
| ES219 | 5471 | M00084893C:A12 | B-30523 |
| ES219 | 5471 | M00084890D:F09 | B-30523 |
| ES219 | 5471 | M00084904A:D03 | B-30523 |
| ES219 | 5471 | M00085029A:C02 | B-30523 |
| ES219 | 5471 | M00084963D:D07 | B-30523 |
| ES219 | 5471 | M00085147C:A04 | B-30523 |
| ES219 | 5471 | M00085144B:C12 | B-30523 |
| ES219 | 5471 | M00085124B:G05 | B-30523 |
| ES219 | 5471 | M00085702B:G11 | B-30523 |
| ES219 | 5471 | M00085203A:E06 | B-30523 |
| ES219 | 5471 | M00085242A:C06 | B-30523 |
| ES219 | 5471 | M00084980D:H08 | B-30523 |
| ES219 | 5471 | M00085187B:C11 | B-30523 |
| ES219 | 5471 | M00085021C:F06 | B-30523 |
| ES219 | 5471 | M00085182B:E04 | B-30523 |
| ES219 | 5471 | M00084930D:B08 | B-30523 |
| ES219 | 5471 | M00084941B:E07 | B-30523 |
| ES219 | 5471 | M00084424D:G07 | B-30523 |
| ES219 | 5471 | M00084938B:F12 | B-30523 |
| ES219 | 5471 | M00084853D:G03 | B-30523 |
| ES219 | 5471 | M00084878B:B12 | B-30523 |
| ES219 | 5471 | M00084889B:C02 | B-30523 |
| ES219 | 5471 | M00084885D:A12 | B-30523 |
| ES219 | 5471 | M00084845A:E02 | B-30523 |
| ES219 | 5471 | M00084972B:H03 | B-30523 |
| ES219 | 5471 | M00084908A:F03 | B-30523 |
| ES219 | 5471 | M00084975A:G05 | B-30523 |
| ES219 | 5471 | M00084941C:H04 | B-30523 |
| ES219 | 5471 | M00084997D:H09 | B-30523 |
| ES219 | 5471 | M00084491A:E08 | B-30523 |
| ES219 | 5471 | M00083815C:H08 | B-30523 |
| ES219 | 5471 | M00084501A:D06 | B-30523 |
| ES219 | 5471 | M00084558D:G08 | B-30523 |
| ES219 | 5471 | M00084510C:F02 | B-30523 |

Table 15

| Library ID | CMCC Number | CloneId | NRRL Number |
|---|---|---|---|
| ES219 | 5471 | M00084521C:H11 | B-30523 |
| ES219 | 5471 | M00084446A:A05 | B-30523 |
| ES219 | 5471 | M00084458A:G06 | B-30523 |
| ES219 | 5471 | M00084377D:E08 | B-30523 |
| ES219 | 5471 | M00084382A:D06 | B-30523 |
| ES219 | 5471 | M00083816B:D08 | B-30523 |
| ES219 | 5471 | M00084449B:C09 | B-30523 |
| ES219 | 5471 | M00084431C:B02 | B-30523 |
| ES219 | 5471 | M00084463A:B07 | B-30523 |
| ES219 | 5471 | M00084487D:F04 | B-30523 |
| ES219 | 5471 | M00083800C:E07 | B-30523 |
| ES219 | 5471 | M00084468C:E07 | B-30523 |
| ES219 | 5471 | M00084638A:E10 | B-30523 |
| ES219 | 5471 | M00084439B:A08 | B-30523 |
| ES219 | 5471 | M00084479D:E10 | B-30523 |
| ES219 | 5471 | M00084455D:B03 | B-30523 |
| ES219 | 5471 | M00084368D:C02 | B-30523 |
| ES219 | 5471 | M00084642D:E08 | B-30523 |
| ES219 | 5471 | M00084373A:F08 | B-30523 |
| ES219 | 5471 | M00084364C:B06 | B-30523 |
| ES219 | 5471 | M00084521B:E11 | B-30523 |
| ES219 | 5471 | M00084385B:D03 | B-30523 |
| ES219 | 5471 | M00084443C:H06 | B-30523 |
| ES219 | 5471 | M00083803C:F03 | B-30523 |
| ES219 | 5471 | M00084421C:B11 | B-30523 |
| ES219 | 5471 | M00084434B:E06 | B-30523 |
| ES219 | 5471 | M00083820B:C03 | B-30523 |
| ES219 | 5471 | M00084246B:H03 | B-30523 |
| ES219 | 5471 | M00084484C:B11 | B-30523 |
| ES219 | 5471 | M00084410C:F10 | B-30523 |
| ES219 | 5471 | M00083801B:H03 | B-30523 |
| ES219 | 5471 | M00084980C:B07 | B-30523 |
| ES219 | 5471 | M00084499C:C11 | B-30523 |
| ES219 | 5471 | M00084526C:G09 | B-30523 |
| ES219 | 5471 | M00084406C:A01 | B-30523 |
| ES219 | 5471 | M00084380D:B07 | B-30523 |
| ES219 | 5471 | M00084383B:A11 | B-30523 |
| ES219 | 5471 | M00083834C:E02 | B-30523 |
| ES219 | 5471 | M00083839A:H03 | B-30523 |
| ES219 | 5471 | M00084505C:H08 | B-30523 |
| ES219 | 5471 | M00084511D:A02 | B-30523 |
| ES219 | 5471 | M00084494C:C01 | B-30523 |
| ES219 | 5471 | M00084451D:F06 | B-30523 |
| ES219 | 5471 | M00084604A:D02 | B-30523 |
| ES219 | 5471 | M00084771D:G03 | B-30523 |
| ES219 | 5471 | M00084817A:H11 | B-30523 |
| ES219 | 5471 | M00084827D:D04 | B-30523 |
| ES219 | 5471 | M00084843D:C06 | B-30523 |
| ES219 | 5471 | M00084750C:B08 | B-30523 |
| ES219 | 5471 | M00084757A:D01 | B-30523 |
| ES219 | 5471 | M00084771D:A01 | B-30523 |
| ES219 | 5471 | M00084730B:A09 | B-30523 |
| ES219 | 5471 | M00084826B:E11 | B-30523 |

Table 15

| Library ID | CMCC Number | CloneId | NRRL Number |
|---|---|---|---|
| ES219 | 5471 | M00084595C:C07 | B-30523 |
| ES219 | 5471 | M00084724A:C02 | B-30523 |
| ES219 | 5471 | M00084833A:G07 | B-30523 |
| ES219 | 5471 | M00084600D:B10 | B-30523 |
| ES219 | 5471 | M00084634C:H02 | B-30523 |
| ES219 | 5471 | M00084614D:A08 | B-30523 |
| ES219 | 5471 | M00084620B:F05 | B-30523 |
| ES219 | 5471 | M00084607A:B03 | B-30523 |
| ES219 | 5471 | M00084633A:B12 | B-30523 |
| ES219 | 5471 | M00084597A:F06 | B-30523 |
| ES219 | 5471 | M00084575A:A11 | B-30523 |
| ES219 | 5471 | M00084547B:B10 | B-30523 |
| ES219 | 5471 | M00084525A:E08 | B-30523 |
| ES219 | 5471 | M00084578B:E12 | B-30523 |
| ES219 | 5471 | M00084669A:A05 | B-30523 |
| ES219 | 5471 | M00084419C:A09 | B-30523 |
| ES219 | 5471 | M00084769C:H03 | B-30523 |
| ES219 | 5471 | M00085007A:B03 | B-30523 |
| ES219 | 5471 | M00084865D:B04 | B-30523 |
| ES219 | 5471 | M00084743D:G01 | B-30523 |
| ES219 | 5471 | M00084770B:G12 | B-30523 |
| ES219 | 5471 | M00084584B:A02 | B-30523 |
| ES219 | 5471 | M00084647C:E12 | B-30523 |
| ES219 | 5471 | M00084766D:F12 | B-30523 |
| ES219 | 5471 | M00084648D:F05 | B-30523 |
| ES219 | 5471 | M00084843A:D06 | B-30523 |
| ES219 | 5471 | M00084709C:B02 | B-30523 |
| ES219 | 5471 | M00084834B:G02 | B-30523 |
| ES219 | 5471 | M00084718D:C04 | B-30523 |
| ES219 | 5471 | M00084702B:C12 | B-30523 |
| ES219 | 5471 | M00084645D:G02 | B-30523 |
| ES219 | 5471 | M00084849B:F11 | B-30523 |
| ES219 | 5471 | M00084859C:H05 | B-30523 |
| ES219 | 5471 | M00084850D:H02 | B-30523 |
| ES219 | 5471 | M00084857B:A09 | B-30523 |
| ES219 | 5471 | M00084867A:C11 | B-30523 |
| ES219 | 5471 | M00084823A:H01 | B-30523 |
| ES219 | 5471 | M00084756B:H01 | B-30523 |
| ES219 | 5471 | M00084700D:E09 | B-30523 |
| ES219 | 5471 | M00085010C:H01 | B-30523 |
| ES219 | 5471 | M00085060B:C05 | B-30523 |
| ES219 | 5471 | M00085012C:A08 | B-30523 |
| ES219 | 5471 | M00085047D:F03 | B-30523 |
| ES219 | 5471 | M00085049B:E03 | B-30523 |
| ES219 | 5471 | M00085051C:A01 | B-30523 |
| ES219 | 5471 | M00085050A:E11 | B-30523 |
| ES219 | 5471 | M00085676C:C04 | B-30523 |
| ES219 | 5471 | M00085121A:D10 | B-30523 |
| ES219 | 5471 | M00085166D:C10 | B-30523 |
| ES219 | 5471 | M00084992D:B02 | B-30523 |
| ES219 | 5471 | M00085148B:H01 | B-30523 |
| ES219 | 5471 | M00085123B:C04 | B-30523 |
| ES219 | 5471 | M00085173B:A08 | B-30523 |

Table 15

| Library ID | CMCC Number | CloneId | NRRL Number |
|---|---|---|---|
| ES219 | 5471 | M00085172C:F06 | B-30523 |
| ES219 | 5471 | M00084937D:B04 | B-30523 |
| ES219 | 5471 | M00085026D:A01 | B-30523 |
| ES219 | 5471 | M00084994D:F11 | B-30523 |
| ES219 | 5471 | M00085190C:D10 | B-30523 |
| ES219 | 5471 | M00085194D:F04 | B-30523 |
| ES219 | 5471 | M00085222D:D07 | B-30523 |
| ES219 | 5471 | M00085223A:G01 | B-30523 |
| ES219 | 5471 | M00084740C:B08 | B-30523 |
| ES219 | 5471 | M00085056A:G12 | B-30523 |
| ES219 | 5471 | M00084671A:C12 | B-30523 |
| ES219 | 5471 | M00084571C:D05 | B-30523 |
| ES219 | 5471 | M00084587B:H07 | B-30523 |
| ES219 | 5471 | M00084582C:H03 | B-30523 |
| ES219 | 5471 | M00084618C:A03 | B-30523 |
| ES219 | 5471 | M00084687A:A03 | B-30523 |
| ES219 | 5471 | M00085038A:C06 | B-30523 |
| ES219 | 5471 | M00084722A:H12 | B-30523 |
| ES219 | 5471 | M00084676B:E02 | B-30523 |
| ES219 | 5471 | M00084615D:H12 | B-30523 |
| ES219 | 5471 | M00084659C:G05 | B-30523 |
| ES219 | 5471 | M00084536B:A03 | B-30523 |
| ES219 | 5471 | M00084929C:B02 | B-30523 |
| ES219 | 5471 | M00084652D:G11 | B-30523 |
| ES219 | 5471 | M00084611B:A11 | B-30523 |
| ES219 | 5471 | M00084530D:G07 | B-30523 |
| ES219 | 5471 | M00084527C:H07 | B-30523 |
| ES219 | 5471 | M00084545C:C05 | B-30523 |
| ES219 | 5471 | M00084535D:C12 | B-30523 |
| ES219 | 5471 | M00084684C:D02 | B-30523 |
| ES219 | 5471 | M00084679D:G12 | B-30523 |
| ES219 | 5471 | M00084734A:E04 | B-30523 |
| ES219 | 5471 | M00084696D:H04 | B-30523 |
| ES220 | 5472 | M00084724D:F04 | B-30524 |
| ES220 | 5472 | M00084559B:F10 | B-30524 |
| ES220 | 5472 | M00084707D:H03 | B-30524 |
| ES220 | 5472 | M00084525D:H01 | B-30524 |
| ES220 | 5472 | M00084578C:G09 | B-30524 |
| ES220 | 5472 | M00084710B:G07 | B-30524 |
| ES220 | 5472 | M00084537B:C05 | B-30524 |
| ES220 | 5472 | M00084560A:G08 | B-30524 |
| ES220 | 5472 | M00085129B:C02 | B-30524 |
| ES220 | 5472 | M00084620D:E05 | B-30524 |
| ES220 | 5472 | M00084576A:E12 | B-30524 |
| ES220 | 5472 | M00084720A:A01 | B-30524 |
| ES220 | 5472 | M00084654A:E04 | B-30524 |
| ES220 | 5472 | M00084596D:E10 | B-30524 |
| ES220 | 5472 | M00084646A:D02 | B-30524 |
| ES220 | 5472 | M00084572D:F07 | B-30524 |
| ES220 | 5472 | M00084620A:E08 | B-30524 |
| ES220 | 5472 | M00084553B:F04 | B-30524 |
| ES220 | 5472 | M00084614D:B07 | B-30524 |
| ES220 | 5472 | M00084604D:D08 | B-30524 |

**Table 15**

| Library ID | CMCC Number | CloneId | NRRL Number |
|---|---|---|---|
| ES220 | 5472 | M00084722D:A03 | B-30524 |
| ES220 | 5472 | M00084958C:B03 | B-30524 |
| ES220 | 5472 | M00084523C:A05 | B-30524 |
| ES220 | 5472 | M00085166C:A08 | B-30524 |
| ES220 | 5472 | M00084467A:D06 | B-30524 |
| ES220 | 5472 | M00084890C:A06 | B-30524 |
| ES220 | 5472 | M00084609C:F10 | B-30524 |
| ES220 | 5472 | M00084413C:A11 | B-30524 |
| ES220 | 5472 | M00084834A:A03 | B-30524 |
| ES220 | 5472 | M00085172A:G05 | B-30524 |
| ES220 | 5472 | M00085146B:C01 | B-30524 |
| ES220 | 5472 | M00085038A:B10 | B-30524 |
| ES220 | 5472 | M00084246A:D03 | B-30524 |
| ES220 | 5472 | M00084967B:D09 | B-30524 |
| ES220 | 5472 | M00085035D:E04 | B-30524 |
| ES220 | 5472 | M00084736B:H03 | B-30524 |
| ES220 | 5472 | M00085025A:D11 | B-30524 |
| ES220 | 5472 | M00084900C:A04 | B-30524 |
| ES220 | 5472 | M00085127C:C03 | B-30524 |
| ES220 | 5472 | M00084424A:G07 | B-30524 |
| ES220 | 5472 | M00085131D:A06 | B-30524 |
| ES220 | 5472 | M00084987B:H12 | B-30524 |
| ES220 | 5472 | M00084967C:D10 | B-30524 |
| ES220 | 5472 | M00084420A:G02 | B-30524 |
| ES220 | 5472 | M00084452B:F07 | B-30524 |
| ES220 | 5472 | M00084705C:D01 | B-30524 |
| ES220 | 5472 | M00085156A:G04 | B-30524 |
| ES220 | 5472 | M00084447D:F03 | B-30524 |
| ES220 | 5472 | M00084495B:C11 | B-30524 |
| ES220 | 5472 | M00084745A:A08 | B-30524 |
| ES220 | 5472 | M00084458B:G05 | B-30524 |
| ES220 | 5472 | M00084449C:C01 | B-30524 |
| ES220 | 5472 | M00084867B:A03 | B-30524 |
| ES220 | 5472 | M00084680A:F08 | B-30524 |
| ES220 | 5472 | M00084585B:D06 | B-30524 |
| ES220 | 5472 | M00084835D:H03 | B-30524 |
| ES220 | 5472 | M00084685C:B12 | B-30524 |
| ES220 | 5472 | M00084500C:D01 | B-30524 |
| ES220 | 5472 | M00084469A:C09 | B-30524 |
| ES220 | 5472 | M00084381C:A05 | B-30524 |
| ES220 | 5472 | M00084477C:C07 | B-30524 |
| ES220 | 5472 | M00084647C:A05 | B-30524 |
| ES220 | 5472 | M00084687C:F12 | B-30524 |
| ES220 | 5472 | M00084756C:H01 | B-30524 |
| ES220 | 5472 | M00084565D:F08 | B-30524 |
| ES220 | 5472 | M00084560B:F12 | B-30524 |
| ES220 | 5472 | M00084640D:A08 | B-30524 |
| ES220 | 5472 | M00084443A:E10 | B-30524 |
| ES220 | 5472 | M00084521A:E11 | B-30524 |
| ES220 | 5472 | M00085019C:D05 | B-30524 |
| ES220 | 5472 | M00084587C:A07 | B-30524 |
| ES220 | 5472 | M00084616A:G03 | B-30524 |
| ES220 | 5472 | M00084732B:A04 | B-30524 |

**Table 15**

| Library ID | CMCC Number | CloneId | NRRL Number |
|---|---|---|---|
| ES220 | 5472 | M00084666A:C04 | B-30524 |
| ES220 | 5472 | M00084633A:H05 | B-30524 |
| ES220 | 5472 | M00084510C:F05 | B-30524 |
| ES220 | 5472 | M00084648B:F06 | B-30524 |
| ES220 | 5472 | M00084700D:H04 | B-30524 |
| ES220 | 5472 | M00084506C:A05 | B-30524 |
| ES220 | 5472 | M00084475C:G11 | B-30524 |
| ES220 | 5472 | M00084673B:H11 | B-30524 |
| ES220 | 5472 | M00084595D:D08 | B-30524 |
| ES220 | 5472 | M00084636C:A06 | B-30524 |
| ES220 | 5472 | M00084612C:B01 | B-30524 |
| ES220 | 5472 | M00084644A:H05 | B-30524 |
| ES220 | 5472 | M00084602D:B09 | B-30524 |
| ES220 | 5472 | M00084584B:G07 | B-30524 |
| ES220 | 5472 | M00084678C:C11 | B-30524 |
| ES220 | 5472 | M00084546C:C06 | B-30524 |
| ES220 | 5472 | M00084755A:D02 | B-30524 |
| ES220 | 5472 | M00084536D:F07 | B-30524 |
| ES220 | 5472 | M00084699A:G05 | B-30524 |
| ES220 | 5472 | M00084438D:H04 | B-30524 |
| ES220 | 5472 | M00084766B:F02 | B-30524 |
| ES220 | 5472 | M00084703B:D09 | B-30524 |
| ES220 | 5472 | M00084856B:D03 | B-30524 |
| ES220 | 5472 | M00084857A:G05 | B-30524 |
| ES220 | 5472 | M00084868B:D01 | B-30524 |
| ES220 | 5472 | M00084823D:E05 | B-30524 |
| ES220 | 5472 | M00084485C:B04 | B-30524 |
| ES220 | 5472 | M00084910D:E07 | B-30524 |
| ES220 | 5472 | M00084996B:D08 | B-30524 |
| ES220 | 5472 | M00084487D:F07 | B-30524 |
| ES220 | 5472 | M00084824C:C10 | B-30524 |
| ES220 | 5472 | M00084949B:B12 | B-30524 |
| ES220 | 5472 | M00084746B:B04 | B-30524 |
| ES220 | 5472 | M00084944D:E05 | B-30524 |
| ES220 | 5472 | M00084851C:F10 | B-30524 |
| ES220 | 5472 | M00084849A:H08 | B-30524 |
| ES220 | 5472 | M00084843A:G01 | B-30524 |
| ES220 | 5472 | M00084921C:E04 | B-30524 |
| ES220 | 5472 | M00084742A:F07 | B-30524 |
| ES220 | 5472 | M00083799D:F10 | B-30524 |
| ES220 | 5472 | M00084760D:D09 | B-30524 |
| ES220 | 5472 | M00084845C:H05 | B-30524 |
| ES220 | 5472 | M00084927A:C01 | B-30524 |
| ES220 | 5472 | M00084935B:E10 | B-30524 |
| ES220 | 5472 | M00084974D:F11 | B-30524 |
| ES220 | 5472 | M00084935C:E07 | B-30524 |
| ES220 | 5472 | M00084503D:G10 | B-30524 |
| ES220 | 5472 | M00084907C:C01 | B-30524 |
| ES220 | 5472 | M00084893C:B01 | B-30524 |
| ES220 | 5472 | M00083803B:F11 | B-30524 |
| ES220 | 5472 | M00084945A:D10 | B-30524 |
| ES220 | 5472 | M00084765B:A10 | B-30524 |
| ES220 | 5472 | M00084455D:G03 | B-30524 |

Table 15

| Library ID | CMCC Number | CloneId | NRRL Number |
|---|---|---|---|
| ES220 | 5472 | M00084874D:E03 | B-30524 |
| ES220 | 5472 | M00084889C:A04 | B-30524 |
| ES220 | 5472 | M00084846B:H07 | B-30524 |
| ES220 | 5472 | M00084967B:B10 | B-30524 |
| ES220 | 5472 | M00084838A:F12 | B-30524 |
| ES220 | 5472 | M00084885A:C01 | B-30524 |
| ES220 | 5472 | M00084823A:H06 | B-30524 |
| ES220 | 5472 | M00084958B:E10 | B-30524 |
| ES220 | 5472 | M00084399B:E05 | B-30524 |
| ES220 | 5472 | M00084880B:D03 | B-30524 |
| ES220 | 5472 | M00084877D:G07 | B-30524 |
| ES220 | 5472 | M00084406B:C03 | B-30524 |
| ES220 | 5472 | M00084856B:A12 | B-30524 |
| ES220 | 5472 | M00084888D:A11 | B-30524 |
| ES220 | 5472 | M00083831D:H11 | B-30524 |
| ES220 | 5472 | M00084481D:C06 | B-30524 |
| ES220 | 5472 | M00083834B:F09 | B-30524 |
| ES220 | 5472 | M00084707D:B08 | B-30524 |
| ES220 | 5472 | M00084976C:C12 | B-30524 |
| ES220 | 5472 | M00085201C:C12 | B-30524 |
| ES220 | 5472 | M00084379D:A05 | B-30524 |
| ES220 | 5472 | M00084392C:G06 | B-30524 |
| ES220 | 5472 | M00084492B:F03 | B-30524 |
| ES220 | 5472 | M00085697B:G05 | B-30524 |
| ES220 | 5472 | M00085683B:B10 | B-30524 |
| ES220 | 5472 | M00084988B:B08 | B-30524 |
| ES220 | 5472 | M00084969C:H11 | B-30524 |
| ES220 | 5472 | M00084988C:G03 | B-30524 |
| ES220 | 5472 | M00085123A:E07 | B-30524 |
| ES220 | 5472 | M00084988C:A04 | B-30524 |
| ES220 | 5472 | M00084363A:C02 | B-30524 |
| ES220 | 5472 | M00084975A:G05 | B-30524 |
| ES220 | 5472 | M00084431C:G08 | B-30524 |
| ES220 | 5472 | M00084972B:H03 | B-30524 |
| ES220 | 5472 | M00084376A:E06 | B-30524 |
| ES220 | 5472 | M00084859C:D09 | B-30524 |
| ES220 | 5472 | M00084957B:H07 | B-30524 |
| ES220 | 5472 | M00085053D:D04 | B-30524 |
| ES220 | 5472 | M00084425A:A01 | B-30524 |
| ES220 | 5472 | M00084367D:E06 | B-30524 |
| ES220 | 5472 | M00084938B:A11 | B-30524 |
| ES220 | 5472 | M00085051A:G03 | B-30524 |
| ES220 | 5472 | M00083817B:G09 | B-30524 |
| ES220 | 5472 | M00085229B:C10 | B-30524 |
| ES220 | 5472 | M00085178D:F01 | B-30524 |
| ES220 | 5472 | M00084980D:D02 | B-30524 |
| ES220 | 5472 | M00085228B:C10 | B-30524 |
| ES220 | 5472 | M00085243A:D07 | B-30524 |
| ES220 | 5472 | M00085031B:E03 | B-30524 |
| ES220 | 5472 | M00085164C:G05 | B-30524 |
| ES220 | 5472 | M00085031C:D05 | B-30524 |
| ES220 | 5472 | M00084251D:C05 | B-30524 |
| ES220 | 5472 | M00085027A:C02 | B-30524 |

Table 15

| Library ID | CMCC Number | CloneId | NRRL Number |
|---|---|---|---|
| ES220 | 5472 | M00084248D:H09 | B-30524 |
| ES220 | 5472 | M00085209C:F11 | B-30524 |
| ES220 | 5472 | M00084368D:D03 | B-30524 |
| ES220 | 5472 | M00083818A:E09 | B-30524 |
| ES220 | 5472 | M00084980C:E06 | B-30524 |
| ES220 | 5472 | M00084248B:C06 | B-30524 |
| ES220 | 5472 | M00085244C:D03 | B-30524 |
| ES220 | 5472 | M00084987A:D09 | B-30524 |
| ES220 | 5472 | M00084994A:H04 | B-30524 |
| ES220 | 5472 | M00084970D:E08 | B-30524 |
| ES220 | 5472 | M00085038D:D10 | B-30524 |
| ES220 | 5472 | M00085035B:C12 | B-30524 |
| ES220 | 5472 | M00085184D:B08 | B-30524 |
| ES221 | 5473 | M00084666C:A06 | B-30525 |
| ES221 | 5473 | M00084657C:E01 | B-30525 |
| ES221 | 5473 | M00084540B:B08 | B-30525 |
| ES221 | 5473 | M00084415C:C05 | B-30525 |
| ES221 | 5473 | M00084812A:C02 | B-30525 |
| ES221 | 5473 | M00084396B:B03 | B-30525 |
| ES221 | 5473 | M00084844B:H08 | B-30525 |
| ES221 | 5473 | M00084877D:H09 | B-30525 |
| ES221 | 5473 | M00084925C:G01 | B-30525 |
| ES221 | 5473 | M00084970A:C11 | B-30525 |
| ES221 | 5473 | M00084961C:F01 | B-30525 |
| ES221 | 5473 | M00084391B:D06 | B-30525 |
| ES221 | 5473 | M00084694D:F04 | B-30525 |
| ES221 | 5473 | M00084698B:D02 | B-30525 |
| ES221 | 5473 | M00084388A:G03 | B-30525 |
| ES221 | 5473 | M00084973A:C01 | B-30525 |
| ES221 | 5473 | M00084423C:G11 | B-30525 |
| ES221 | 5473 | M00084497D:D03 | B-30525 |
| ES221 | 5473 | M00084889D:G06 | B-30525 |
| ES221 | 5473 | M00084959B:C07 | B-30525 |
| ES221 | 5473 | M00084432B:C05 | B-30525 |
| ES221 | 5473 | M00084489A:D12 | B-30525 |
| ES221 | 5473 | M00084748A:D09 | B-30525 |
| ES221 | 5473 | M00084962C:F10 | B-30525 |
| ES221 | 5473 | M00084767B:D10 | B-30525 |
| ES221 | 5473 | M00084711B:A05 | B-30525 |
| ES221 | 5473 | M00084743A:E03 | B-30525 |
| ES221 | 5473 | M00084466B:E01 | B-30525 |
| ES221 | 5473 | M00084450C:A09 | B-30525 |
| ES221 | 5473 | M00084492C:B05 | B-30525 |
| ES221 | 5473 | M00084487B:A06 | B-30525 |
| ES221 | 5473 | M00084480B:A05 | B-30525 |
| ES221 | 5473 | M00084764D:G08 | B-30525 |
| ES221 | 5473 | M00084743D:H04 | B-30525 |
| ES221 | 5473 | M00084891D:A02 | B-30525 |
| ES221 | 5473 | M00084822C:D06 | B-30525 |
| ES221 | 5473 | M00084853D:A12 | B-30525 |
| ES221 | 5473 | M00084822B:G11 | B-30525 |
| ES221 | 5473 | M00084756D:C04 | B-30525 |
| ES221 | 5473 | M00084839C:B09 | B-30525 |

Table 15

| Library ID | CMCC Number | CloneId | NRRL Number |
|---|---|---|---|
| ES221 | 5473 | M00084767D:B04 | B-30525 |
| ES221 | 5473 | M00084703A:E04 | B-30525 |
| ES221 | 5473 | M00084853A:F08 | B-30525 |
| ES221 | 5473 | M00084956C:G09 | B-30525 |
| ES221 | 5473 | M00084908C:F07 | B-30525 |
| ES221 | 5473 | M00084902B:A10 | B-30525 |
| ES221 | 5473 | M00084833D:B04 | B-30525 |
| ES221 | 5473 | M00085023D:E11 | B-30525 |
| ES221 | 5473 | M00085151A:B04 | B-30525 |
| ES221 | 5473 | M00085039D:F09 | B-30525 |
| ES221 | 5473 | M00085169A:H12 | B-30525 |
| ES221 | 5473 | M00085052B:E04 | B-30525 |
| ES221 | 5473 | M00085171D:F05 | B-30525 |
| ES221 | 5473 | M00085050A:B06 | B-30525 |
| ES221 | 5473 | M00085155B:F10 | B-30525 |
| ES221 | 5473 | M00085123C:G11 | B-30525 |
| ES221 | 5473 | M00085182B:H10 | B-30525 |
| ES221 | 5473 | M00084675A:E02 | B-30525 |
| ES221 | 5473 | M00085248B:G12 | B-30525 |
| ES221 | 5473 | M00084731C:G07 | B-30525 |
| ES221 | 5473 | M00085701A:A09 | B-30525 |
| ES221 | 5473 | M00085246B:G12 | B-30525 |
| ES221 | 5473 | M00084967C:D12 | B-30525 |
| ES221 | 5473 | M00085190B:C09 | B-30525 |
| ES221 | 5473 | M00085167C:D06 | B-30525 |
| ES221 | 5473 | M00085705A:E01 | B-30525 |
| ES221 | 5473 | M00085214D:G01 | B-30525 |
| ES221 | 5473 | M00084755D:E06 | B-30525 |
| ES221 | 5473 | M00084630D:F09 | B-30525 |
| ES221 | 5473 | M00085191A:B03 | B-30525 |
| ES221 | 5473 | M00085143C:D05 | B-30525 |
| ES221 | 5473 | M00084886A:C06 | B-30525 |
| ES221 | 5473 | M00083803B:F12 | B-30525 |
| ES221 | 5473 | M00084949B:H11 | B-30525 |
| ES221 | 5473 | M00084701C:E08 | B-30525 |
| ES221 | 5473 | M00084945A:H10 | B-30525 |
| ES221 | 5473 | M00084667C:A03 | B-30525 |
| ES221 | 5473 | M00084953D:D03 | B-30525 |
| ES221 | 5473 | M00084539D:D11 | B-30525 |
| ES221 | 5473 | M00084737A:C09 | B-30525 |
| ES221 | 5473 | M00084968C:D10 | B-30525 |
| ES221 | 5473 | M00084670B:A09 | B-30525 |
| ES221 | 5473 | M00085167A:G02 | B-30525 |
| ES221 | 5473 | M00084554C:D05 | B-30525 |
| ES221 | 5473 | M00085145C:D02 | B-30525 |
| ES221 | 5473 | M00084722D:G04 | B-30525 |
| ES221 | 5473 | M00084721C:F09 | B-30525 |
| ES221 | 5473 | M00084866B:A03 | B-30525 |
| ES221 | 5473 | M00084727A:A02 | B-30525 |
| ES221 | 5473 | M00084407A:H09 | B-30525 |
| ES221 | 5473 | M00084855D:H05 | B-30525 |
| ES221 | 5473 | M00084403D:D04 | B-30525 |
| ES221 | 5473 | M00085144D:G03 | B-30525 |

**Table 15**

| Library ID | CMCC Number | CloneId | NRRL Number |
|---|---|---|---|
| ES221 | 5473 | M00084880D:A10 | B-30525 |
| ES221 | 5473 | M00084958C:B03 | B-30525 |
| ES221 | 5473 | M00084888C:D12 | B-30525 |
| ES221 | 5473 | M00084587C:G07 | B-30525 |
| ES221 | 5473 | M00083844C:C04 | B-30525 |
| ES221 | 5473 | M00084647D:C05 | B-30525 |
| ES221 | 5473 | M00084528C:F06 | B-30525 |
| ES221 | 5473 | M00084857D:A11 | B-30525 |
| ES221 | 5473 | M00084385A:D02 | B-30525 |
| ES221 | 5473 | M00084561C:D07 | B-30525 |
| ES221 | 5473 | M00084994D:H04 | B-30525 |
| ES221 | 5473 | M00084448B:D11 | B-30525 |
| ES221 | 5473 | M00085006D:C10 | B-30525 |
| ES221 | 5473 | M00084580B:B05 | B-30525 |
| ES221 | 5473 | M00083814D:A10 | B-30525 |
| ES221 | 5473 | M00084970C:G03 | B-30525 |
| ES221 | 5473 | M00084372D:H11 | B-30525 |
| ES221 | 5473 | M00084377B:E11 | B-30525 |
| ES221 | 5473 | M00085230B:G08 | B-30525 |
| ES221 | 5473 | M00084584B:F09 | B-30525 |
| ES221 | 5473 | M00084584B:H12 | B-30525 |
| ES221 | 5473 | M00085249C:C11 | B-30525 |
| ES221 | 5473 | M00084441B:E05 | B-30525 |
| ES221 | 5473 | M00083841A:G01 | B-30525 |
| ES221 | 5473 | M00085006D:C04 | B-30525 |
| ES221 | 5473 | M00084686B:B04 | B-30525 |
| ES221 | 5473 | M00084998A:C12 | B-30525 |
| ES221 | 5473 | M00085034B:E11 | B-30525 |
| ES221 | 5473 | M00084683B:A01 | B-30525 |
| ES221 | 5473 | M00084613A:A01 | B-30525 |
| ES221 | 5473 | M00084633B:A06 | B-30525 |
| ES221 | 5473 | M00085032C:F04 | B-30525 |
| ES221 | 5473 | M00085022B:F05 | B-30525 |
| ES221 | 5473 | M00084509A:E10 | B-30525 |
| ES221 | 5473 | M00084400A:B09 | B-30525 |
| ES221 | 5473 | M00084677C:F03 | B-30525 |
| ES221 | 5473 | M00084427B:D01 | B-30525 |
| ES221 | 5473 | M00083844B:C04 | B-30525 |
| ES221 | 5473 | M00084598D:H05 | B-30525 |
| ES221 | 5473 | M00084443B:C02 | B-30525 |
| ES221 | 5473 | M00084514A:A03 | B-30525 |
| ES221 | 5473 | M00084560C:G05 | B-30525 |
| ES221 | 5473 | M00084504C:F05 | B-30525 |
| ES221 | 5473 | M00084517C:D06 | B-30525 |
| ES221 | 5473 | M00084420C:D03 | B-30525 |
| ES221 | 5473 | M00084524D:D02 | B-30525 |
| ES221 | 5473 | M00084499D:A10 | B-30525 |
| ES221 | 5473 | M00085022B:B03 | B-30525 |
| ES221 | 5473 | M00084958B:E10 | B-30525 |
| ES221 | 5473 | M00084513C:C10 | B-30525 |
| ES221 | 5473 | M00084595B:C08 | B-30525 |
| ES221 | 5473 | M00083804A:H12 | B-30525 |
| ES221 | 5473 | M00084859D:B03 | B-30525 |

Table 15

| Library ID | CMCC Number | CloneId | NRRL Number |
|---|---|---|---|
| ES221 | 5473 | M00084641B:F08 | B-30525 |
| ES221 | 5473 | M00084844A:E10 | B-30525 |
| ES221 | 5473 | M00083838A:E05 | B-30525 |
| ES221 | 5473 | M00083849C:F11 | B-30525 |
| ES221 | 5473 | M00084461C:D06 | B-30525 |
| ES221 | 5473 | M00084810D:B10 | B-30525 |
| ES221 | 5473 | M00085047D:F08 | B-30525 |
| ES221 | 5473 | M00084912D:G06 | B-30525 |
| ES221 | 5473 | M00084645C:F07 | B-30525 |
| ES221 | 5473 | M00084912D:A09 | B-30525 |
| ES221 | 5473 | M00084862B:B01 | B-30525 |
| ES221 | 5473 | M00084938C:G06 | B-30525 |
| ES221 | 5473 | M00084534B:E12 | B-30525 |
| ES221 | 5473 | M00084909C:G02 | B-30525 |
| ES221 | 5473 | M00084973A:A01 | B-30525 |
| ES221 | 5473 | M00084651B:G10 | B-30525 |
| ES221 | 5473 | M00084925A:B08 | B-30525 |
| ES221 | 5473 | M00084568D:A02 | B-30525 |
| ES221 | 5473 | M00084456A:H04 | B-30525 |
| ES221 | 5473 | M00084988C:B01 | B-30525 |
| ES221 | 5473 | M00084842C:B07 | B-30525 |
| ES221 | 5473 | M00084708A:A11 | B-30525 |
| ES221 | 5473 | M00084602C:E04 | B-30525 |
| ES221 | 5473 | M00084757B:F11 | B-30525 |
| ES221 | 5473 | M00084483A:C06 | B-30525 |
| ES221 | 5473 | M00084605B:H04 | B-30525 |
| ES221 | 5473 | M00083812C:G02 | B-30525 |
| ES221 | 5473 | M00084610D:H04 | B-30525 |
| ES221 | 5473 | M00085056D:B12 | B-30525 |
| ES221 | 5473 | M00085017C:A11 | B-30525 |
| ES221 | 5473 | M00084573A:A10 | B-30525 |
| ES221 | 5473 | M00084637B:E01 | B-30525 |
| ES221 | 5473 | M00085056B:B06 | B-30525 |
| ES221 | 5473 | M00084510C:H01 | B-30525 |
| ES221 | 5473 | M00084577B:C08 | B-30525 |
| ES221 | 5473 | M00084646B:B03 | B-30525 |
| ES221 | 5473 | M00084844C:F04 | B-30525 |
| ES221 | 5473 | M00084894B:F11 | B-30525 |
| ES221 | 5473 | M00084930B:E12 | B-30525 |
| ES221 | 5473 | M00084469B:F08 | B-30525 |
| ES221 | 5473 | M00084569D:B04 | B-30525 |
| ES221 | 5473 | M00084453D:B12 | B-30525 |
| ES221 | 5473 | M00083844A:E12 | B-30525 |
| ES221 | 5473 | M00085009D:A02 | B-30525 |
| ES221 | 5473 | M00084619A:E04 | B-30525 |
| ES221 | 5473 | M00085006C:C07 | B-30525 |
| ES222 | 5474 | M00084459A:F10 | B-30526 |
| ES222 | 5474 | M00084721B:C11 | B-30526 |
| ES222 | 5474 | M00084454A:G08 | B-30526 |
| ES222 | 5474 | M00084460D:B04 | B-30526 |
| ES222 | 5474 | M00084723D:G09 | B-30526 |
| ES222 | 5474 | M00084704A:C12 | B-30526 |
| ES222 | 5474 | M00084487C:H06 | B-30526 |

Table 15

| Library ID | CMCC Number | CloneId | NRRL Number |
|---|---|---|---|
| ES222 | 5474 | M00084867C:G02 | B-30526 |
| ES222 | 5474 | M00084475B:D03 | B-30526 |
| ES222 | 5474 | M00084490A:C12 | B-30526 |
| ES222 | 5474 | M00084865D:G02 | B-30526 |
| ES222 | 5474 | M00084876D:A06 | B-30526 |
| ES222 | 5474 | M00084553D:G05 | B-30526 |
| ES222 | 5474 | M00084558D:A04 | B-30526 |
| ES222 | 5474 | M00084645B:A06 | B-30526 |
| ES222 | 5474 | M00084747D:G02 | B-30526 |
| ES222 | 5474 | M00084884D:D03 | B-30526 |
| ES222 | 5474 | M00084700A:C10 | B-30526 |
| ES222 | 5474 | M00084973A:C01 | B-30526 |
| ES222 | 5474 | M00084493A:E03 | B-30526 |
| ES222 | 5474 | M00084497B:C12 | B-30526 |
| ES222 | 5474 | M00084500D:B11 | B-30526 |
| ES222 | 5474 | M00084523C:C10 | B-30526 |
| ES222 | 5474 | M00084526D:E09 | B-30526 |
| ES222 | 5474 | M00084923D:B05 | B-30526 |
| ES222 | 5474 | M00084962C:F10 | B-30526 |
| ES222 | 5474 | M00084669C:A10 | B-30526 |
| ES222 | 5474 | M00084444D:F09 | B-30526 |
| ES222 | 5474 | M00084757B:F05 | B-30526 |
| ES222 | 5474 | M00084922A:C08 | B-30526 |
| ES222 | 5474 | M00084960D:D02 | B-30526 |
| ES222 | 5474 | M00084837B:E06 | B-30526 |
| ES222 | 5474 | M00084763D:A04 | B-30526 |
| ES222 | 5474 | M00084651C:H01 | B-30526 |
| ES222 | 5474 | M00084441D:E09 | B-30526 |
| ES222 | 5474 | M00084509D:C02 | B-30526 |
| ES222 | 5474 | M00084510D:D05 | B-30526 |
| ES222 | 5474 | M00084657D:B10 | B-30526 |
| ES222 | 5474 | M00084946D:H05 | B-30526 |
| ES222 | 5474 | M00084506A:E08 | B-30526 |
| ES222 | 5474 | M00084420D:C07 | B-30526 |
| ES222 | 5474 | M00085247A:F05 | B-30526 |
| ES222 | 5474 | M00085142D:F04 | B-30526 |
| ES222 | 5474 | M00085151A:H09 | B-30526 |
| ES222 | 5474 | M00085029D:E12 | B-30526 |
| ES222 | 5474 | M00085141C:G06 | B-30526 |
| ES222 | 5474 | M00085035D:D09 | B-30526 |
| ES222 | 5474 | M00085168D:D04 | B-30526 |
| ES222 | 5474 | M00084995B:B08 | B-30526 |
| ES222 | 5474 | M00085008B:H11 | B-30526 |
| ES222 | 5474 | M00085059B:H11 | B-30526 |
| ES222 | 5474 | M00085125C:H06 | B-30526 |
| ES222 | 5474 | M00084890B:E02 | B-30526 |
| ES222 | 5474 | M00084418D:A04 | B-30526 |
| ES222 | 5474 | M00084961C:A06 | B-30526 |
| ES222 | 5474 | M00084766B:E03 | B-30526 |
| ES222 | 5474 | M00084406A:B03 | B-30526 |
| ES222 | 5474 | M00085686A:C05 | B-30526 |
| ES222 | 5474 | M00085124A:G04 | B-30526 |
| ES222 | 5474 | M00085059B:H07 | B-30526 |

Table 15

| Library ID | CMCC Number | CloneId | NRRL Number |
|---|---|---|---|
| ES222 | 5474 | M00084646B:D07 | B-30526 |
| ES222 | 5474 | M00084246B:D10 | B-30526 |
| ES222 | 5474 | M00084974D:F11 | B-30526 |
| ES222 | 5474 | M00084738B:A09 | B-30526 |
| ES222 | 5474 | M00085015A:C09 | B-30526 |
| ES222 | 5474 | M00083839B:G09 | B-30526 |
| ES222 | 5474 | M00085012C:D06 | B-30526 |
| ES222 | 5474 | M00085058A:H02 | B-30526 |
| ES222 | 5474 | M00085009C:C01 | B-30526 |
| ES222 | 5474 | M00083818C:A02 | B-30526 |
| ES222 | 5474 | M00085007B:C07 | B-30526 |
| ES222 | 5474 | M00085047A:H02 | B-30526 |
| ES222 | 5474 | M00085245C:D07 | B-30526 |
| ES222 | 5474 | M00085032D:C03 | B-30526 |
| ES222 | 5474 | M00085039B:E09 | B-30526 |
| ES222 | 5474 | M00085053C:D07 | B-30526 |
| ES222 | 5474 | M00084364D:F08 | B-30526 |
| ES222 | 5474 | M00085152B:A06 | B-30526 |
| ES222 | 5474 | M00084969C:F03 | B-30526 |
| ES222 | 5474 | M00084896B:G10 | B-30526 |
| ES222 | 5474 | M00084427C:D04 | B-30526 |
| ES222 | 5474 | M00085018C:B09 | B-30526 |
| ES222 | 5474 | M00084668D:D08 | B-30526 |
| ES222 | 5474 | M00085175B:A03 | B-30526 |
| ES222 | 5474 | M00084437C:G05 | B-30526 |
| ES222 | 5474 | M00084967B:B10 | B-30526 |
| ES222 | 5474 | M00084998B:A04 | B-30526 |
| ES222 | 5474 | M00084716D:H03 | B-30526 |
| ES222 | 5474 | M00084708B:A06 | B-30526 |
| ES222 | 5474 | M00084755B:A04 | B-30526 |
| ES222 | 5474 | M00084380C:C09 | B-30526 |
| ES222 | 5474 | M00085013A:E06 | B-30526 |
| ES222 | 5474 | M00084374A:A10 | B-30526 |
| ES222 | 5474 | M00085194C:B12 | B-30526 |
| ES222 | 5474 | M00084971C:G07 | B-30526 |
| ES222 | 5474 | M00084515D:G03 | B-30526 |
| ES222 | 5474 | M00084961D:H03 | B-30526 |
| ES222 | 5474 | M00084908D:B11 | B-30526 |
| ES222 | 5474 | M00084702A:B08 | B-30526 |
| ES222 | 5474 | M00083838C:F07 | B-30526 |
| ES222 | 5474 | M00084390B:H04 | B-30526 |
| ES222 | 5474 | M00084734A:H01 | B-30526 |
| ES222 | 5474 | M00083817B:A11 | B-30526 |
| ES222 | 5474 | M00085176C:B11 | B-30526 |
| ES222 | 5474 | M00084902C:F05 | B-30526 |
| ES222 | 5474 | M00085677A:E02 | B-30526 |
| ES222 | 5474 | M00084948D:B08 | B-30526 |
| ES222 | 5474 | M00085190B:H04 | B-30526 |
| ES222 | 5474 | M00084820D:A03 | B-30526 |
| ES222 | 5474 | M00084479B:E04 | B-30526 |
| ES222 | 5474 | M00084408D:E06 | B-30526 |
| ES222 | 5474 | M00085009B:F10 | B-30526 |
| ES222 | 5474 | M00085697A:F01 | B-30526 |

Table 15

| Library ID | CMCC Number | CloneId | NRRL Number |
|---|---|---|---|
| ES222 | 5474 | M00084423D:B05 | B-30526 |
| ES222 | 5474 | M00084973A:A01 | B-30526 |
| ES222 | 5474 | M00083804B:C03 | B-30526 |
| ES222 | 5474 | M00084841B:H09 | B-30526 |
| ES222 | 5474 | M00084685D:B11 | B-30526 |
| ES222 | 5474 | M00084599D:C02 | B-30526 |
| ES222 | 5474 | M00084573D:G11 | B-30526 |
| ES222 | 5474 | M00084603A:B07 | B-30526 |
| ES222 | 5474 | M00084823D:E06 | B-30526 |
| ES222 | 5474 | M00084565A:D10 | B-30526 |
| ES222 | 5474 | M00084767B:F06 | B-30526 |
| ES222 | 5474 | M00084963D:D07 | B-30526 |
| ES222 | 5474 | M00084611A:A06 | B-30526 |
| ES222 | 5474 | M00084829B:F06 | B-30526 |
| ES222 | 5474 | M00084850C:A11 | B-30526 |
| ES222 | 5474 | M00084540D:B12 | B-30526 |
| ES222 | 5474 | M00084614C:G05 | B-30526 |
| ES222 | 5474 | M00084826B:D12 | B-30526 |
| ES222 | 5474 | M00084605D:G09 | B-30526 |
| ES222 | 5474 | M00084923D:F11 | B-30526 |
| ES222 | 5474 | M00084664D:E05 | B-30526 |
| ES222 | 5474 | M00084533A:C04 | B-30526 |
| ES222 | 5474 | M00084843D:F05 | B-30526 |
| ES222 | 5474 | M00084894A:G09 | B-30526 |
| ES222 | 5474 | M00084913B:F05 | B-30526 |
| ES222 | 5474 | M00083817D:A08 | B-30526 |
| ES222 | 5474 | M00084451D:A03 | B-30526 |
| ES222 | 5474 | M00084675B:A04 | B-30526 |
| ES222 | 5474 | M00084889A:B07 | B-30526 |
| ES222 | 5474 | M00084879A:A04 | B-30526 |
| ES222 | 5474 | M00084638D:A05 | B-30526 |
| ES222 | 5474 | M00084468A:A09 | B-30526 |
| ES222 | 5474 | M00084634A:D01 | B-30526 |
| ES222 | 5474 | M00084577B:D04 | B-30526 |
| ES222 | 5474 | M00084860B:A01 | B-30526 |
| ES222 | 5474 | M00084567B:F03 | B-30526 |
| ES222 | 5474 | M00084619A:G10 | B-30526 |
| ES222 | 5474 | M00084683A:B12 | B-30526 |
| ES222 | 5474 | M00084631D:G01 | B-30526 |
| ES222 | 5474 | M00084520B:A12 | B-30526 |
| ES222 | 5474 | M00084886B:D06 | B-30526 |
| ES222 | 5474 | M00084727A:G09 | B-30526 |
| ES222 | 5474 | M00084393A:G07 | B-30526 |
| ES222 | 5474 | M00084571A:C02 | B-30526 |
| ES222 | 5474 | M00084866C:H04 | B-30526 |
| ES222 | 5474 | M00084449A:D09 | B-30526 |
| ES222 | 5474 | M00084857C:E11 | B-30526 |
| ES222 | 5474 | M00085226C:F08 | B-30526 |
| ES222 | 5474 | M00084392C:D03 | B-30526 |
| ES222 | 5474 | M00084389A:F12 | B-30526 |
| ES222 | 5474 | M00084696C:A07 | B-30526 |
| ES222 | 5474 | M00084397D:A09 | B-30526 |
| ES222 | 5474 | M00085173A:B07 | B-30526 |

Table 15

| Library ID | CMCC Number | CloneId | NRRL Number |
|---|---|---|---|
| ES222 | 5474 | M00084252B:H01 | B-30526 |
| ES222 | 5474 | M00084970C:H09 | B-30526 |
| ES222 | 5474 | M00084648A:F08 | B-30526 |
| ES222 | 5474 | M00085245D:G07 | B-30526 |
| ES222 | 5474 | M00084642C:F10 | B-30526 |
| ES222 | 5474 | M00085220D:E06 | B-30526 |
| ES222 | 5474 | M00084745A:H04 | B-30526 |
| ES222 | 5474 | M00083809B:E08 | B-30526 |
| ES222 | 5474 | M00084940B:F06 | B-30526 |
| ES222 | 5474 | M00084533B:B10 | B-30526 |
| ES222 | 5474 | M00084970D:B01 | B-30526 |
| ES222 | 5474 | M00084583D:H12 | B-30526 |
| ES222 | 5474 | M00084585D:H12 | B-30526 |
| ES222 | 5474 | M00084581B:E06 | B-30526 |
| ES222 | 5474 | M00084588B:D02 | B-30526 |
| ES222 | 5474 | M00084919D:B08 | B-30526 |
| ES222 | 5474 | M00084812A:E05 | B-30526 |
| ES222 | 5474 | M00084768B:E09 | B-30526 |
| ES222 | 5474 | M00084748A:H02 | B-30526 |
| ES222 | 5474 | M00084519B:D01 | B-30526 |
| ES222 | 5474 | M00084926B:C05 | B-30526 |
| ES222 | 5474 | M00084847B:G05 | B-30526 |
| ES222 | 5474 | M00084858C:B01 | B-30526 |
| ES222 | 5474 | M00084483A:E05 | B-30526 |
| ES222 | 5474 | M00084596A:G03 | B-30526 |
| ES222 | 5474 | M00084681B:G11 | B-30526 |
| ES223 | 5475 | M00085368B:A02 | B-30527 |
| ES223 | 5475 | M00085365C:C09 | B-30527 |
| ES223 | 5475 | M00085317B:G09 | B-30527 |
| ES223 | 5475 | M00085732A:B09 | B-30527 |
| ES223 | 5475 | M00085649C:A12 | B-30527 |
| ES223 | 5475 | M00085337C:E09 | B-30527 |
| ES223 | 5475 | M00085520C:D02 | B-30527 |
| ES223 | 5475 | M00085358B:A04 | B-30527 |
| ES223 | 5475 | M00085640A:H11 | B-30527 |
| ES223 | 5475 | M00085344D:B07 | B-30527 |
| ES223 | 5475 | M00085314C:F01 | B-30527 |
| ES223 | 5475 | M00085334A:D10 | B-30527 |
| ES223 | 5475 | M00085262C:E04 | B-30527 |
| ES223 | 5475 | M00083750D:G12 | B-30527 |
| ES223 | 5475 | M00085255B:F11 | B-30527 |
| ES223 | 5475 | M00085701D:A02 | B-30527 |
| ES223 | 5475 | M00086280B:G09 | B-30527 |
| ES223 | 5475 | M00085628C:D08 | B-30527 |
| ES223 | 5475 | M00083726B:E07 | B-30527 |
| ES223 | 5475 | M00086285B:C10 | B-30527 |
| ES223 | 5475 | M00086084D:E12 | B-30527 |
| ES223 | 5475 | M00085446A:D04 | B-30527 |
| ES223 | 5475 | M00085697D:H11 | B-30527 |
| ES223 | 5475 | M00086085A:H03 | B-30527 |
| ES223 | 5475 | M00086057A:F07 | B-30527 |
| ES223 | 5475 | M00086196A:F07 | B-30527 |
| ES223 | 5475 | M00086279A:B07 | B-30527 |

**Table 15**

| Library ID | CMCC Number | CloneId | NRRL Number |
|---|---|---|---|
| ES223 | 5475 | M00086266B:E10 | B-30527 |
| ES223 | 5475 | M00086280A:E02 | B-30527 |
| ES223 | 5475 | M00085309A:D11 | B-30527 |
| ES223 | 5475 | M00086291D:B08 | B-30527 |
| ES223 | 5475 | M00085861C:A03 | B-30527 |
| ES223 | 5475 | M00086247A:G11 | B-30527 |
| ES223 | 5475 | M00085373C:B06 | B-30527 |
| ES223 | 5475 | M00085332D:B03 | B-30527 |
| ES223 | 5475 | M00085632C:A06 | B-30527 |
| ES223 | 5475 | M00085360B:G03 | B-30527 |
| ES223 | 5475 | M00086191A:G09 | B-30527 |
| ES223 | 5475 | M00085473C:B02 | B-30527 |
| ES223 | 5475 | M00085600B:B03 | B-30527 |
| ES223 | 5475 | M00083749D:B09 | B-30527 |
| ES223 | 5475 | M00085301B:C10 | B-30527 |
| ES223 | 5475 | M00085278D:F03 | B-30527 |
| ES223 | 5475 | M00085296C:C09 | B-30527 |
| ES223 | 5475 | M00085435B:C05 | B-30527 |
| ES223 | 5475 | M00085533C:D11 | B-30527 |
| ES223 | 5475 | M00086061B:E02 | B-30527 |
| ES223 | 5475 | M00085284D:A09 | B-30527 |
| ES223 | 5475 | M00085566A:D12 | B-30527 |
| ES223 | 5475 | M00085528A:E02 | B-30527 |
| ES223 | 5475 | M00085336D:B10 | B-30527 |
| ES223 | 5475 | M00085315C:B03 | B-30527 |
| ES223 | 5475 | M00085509A:A02 | B-30527 |
| ES223 | 5475 | M00083698D:E01 | B-30527 |
| ES223 | 5475 | M00083701D:G09 | B-30527 |
| ES223 | 5475 | M00086155A:G12 | B-30527 |
| ES223 | 5475 | M00085293B:B05 | B-30527 |
| ES223 | 5475 | M00085728B:C08 | B-30527 |
| ES223 | 5475 | M00085611B:D03 | B-30527 |
| ES223 | 5475 | M00085592A:G06 | B-30527 |
| ES223 | 5475 | M00085304A:B11 | B-30527 |
| ES223 | 5475 | M00085266D:C09 | B-30527 |
| ES223 | 5475 | M00085335B:D09 | B-30527 |
| ES223 | 5475 | M00085707C:A10 | B-30527 |
| ES223 | 5475 | M00085555D:F08 | B-30527 |
| ES223 | 5475 | M00085588B:G10 | B-30527 |
| ES223 | 5475 | M00085264C:F04 | B-30527 |
| ES223 | 5475 | M00085733D:E05 | B-30527 |
| ES223 | 5475 | M00085647A:C08 | B-30527 |
| ES223 | 5475 | M00083714C:F04 | B-30527 |
| ES223 | 5475 | M00085707A:F01 | B-30527 |
| ES223 | 5475 | M00085548C:D04 | B-30527 |
| ES223 | 5475 | M00083745A:A10 | B-30527 |
| ES223 | 5475 | M00085396B:G04 | B-30527 |
| ES223 | 5475 | M00085449C:D04 | B-30527 |
| ES223 | 5475 | M00083698B:H01 | B-30527 |
| ES223 | 5475 | M00084772C:G12 | B-30527 |
| ES223 | 5475 | M00086126C:D09 | B-30527 |
| ES223 | 5475 | M00085808D:E01 | B-30527 |
| ES223 | 5475 | M00085927A:F06 | B-30527 |

**Table 15**

| Library ID | CMCC Number | CloneId | NRRL Number |
|---|---|---|---|
| ES223 | 5475 | M00085814C:C12 | B-30527 |
| ES223 | 5475 | M00086015A:B03 | B-30527 |
| ES223 | 5475 | M00086146D:A09 | B-30527 |
| ES223 | 5475 | M00085076C:A07 | B-30527 |
| ES223 | 5475 | M00085427C:A04 | B-30527 |
| ES223 | 5475 | M00084774B:C10 | B-30527 |
| ES223 | 5475 | M00085432A:H08 | B-30527 |
| ES223 | 5475 | M00084796D:B01 | B-30527 |
| ES223 | 5475 | M00086127C:C05 | B-30527 |
| ES223 | 5475 | M00086160D:F08 | B-30527 |
| ES223 | 5475 | M00084802A:H09 | B-30527 |
| ES223 | 5475 | M00086081D:H11 | B-30527 |
| ES223 | 5475 | M00086106D:H01 | B-30527 |
| ES223 | 5475 | M00086159A:F05 | B-30527 |
| ES223 | 5475 | M00084782D:H08 | B-30527 |
| ES223 | 5475 | M00085956D:G04 | B-30527 |
| ES223 | 5475 | M00085770C:A12 | B-30527 |
| ES223 | 5475 | M00086008D:F08 | B-30527 |
| ES223 | 5475 | M00086018A:A05 | B-30527 |
| ES223 | 5475 | M00085761A:B03 | B-30527 |
| ES223 | 5475 | M00085751A:A11 | B-30527 |
| ES223 | 5475 | M00085956B:E08 | B-30527 |
| ES223 | 5475 | M00085955C:C03 | B-30527 |
| ES223 | 5475 | M00085904D:D02 | B-30527 |
| ES223 | 5475 | M00085899C:G10 | B-30527 |
| ES223 | 5475 | M00085927C:G10 | B-30527 |
| ES223 | 5475 | M00085896D:A11 | B-30527 |
| ES223 | 5475 | M00085892A:F04 | B-30527 |
| ES223 | 5475 | M00085882B:F11 | B-30527 |
| ES223 | 5475 | M00085419A:G09 | B-30527 |
| ES223 | 5475 | M00085962B:A12 | B-30527 |
| ES223 | 5475 | M00085811B:D12 | B-30527 |
| ES223 | 5475 | M00085986B:H02 | B-30527 |
| ES223 | 5475 | M00085922A:A08 | B-30527 |
| ES223 | 5475 | M00085854C:F04 | B-30527 |
| ES223 | 5475 | M00085835D:F06 | B-30527 |
| ES223 | 5475 | M00086183A:H04 | B-30527 |
| ES223 | 5475 | M00086193A:F04 | B-30527 |
| ES223 | 5475 | M00086197B:A03 | B-30527 |
| ES223 | 5475 | M00086203C:H04 | B-30527 |
| ES223 | 5475 | M00085827D:D01 | B-30527 |
| ES223 | 5475 | M00086097D:D12 | B-30527 |
| ES223 | 5475 | M00086294C:G05 | B-30527 |
| ES223 | 5475 | M00086176C:A06 | B-30527 |
| ES223 | 5475 | M00085825C:D12 | B-30527 |
| ES223 | 5475 | M00085849D:G06 | B-30527 |
| ES223 | 5475 | M00085817C:C10 | B-30527 |
| ES223 | 5475 | M00085807D:G11 | B-30527 |
| ES223 | 5475 | M00085839B:B12 | B-30527 |
| ES223 | 5475 | M00085750A:G03 | B-30527 |
| ES223 | 5475 | M00086248A:H09 | B-30527 |
| ES223 | 5475 | M00085964A:B11 | B-30527 |
| ES223 | 5475 | M00086003D:G08 | B-30527 |

**Table 15**

| Library ID | CMCC Number | CloneId | NRRL Number |
|---|---|---|---|
| ES223 | 5475 | M00085819D:C02 | B-30527 |
| ES223 | 5475 | M00085066B:D12 | B-30527 |
| ES223 | 5475 | M00084775C:E05 | B-30527 |
| ES223 | 5475 | M00085083A:E04 | B-30527 |
| ES223 | 5475 | M00085090C:C09 | B-30527 |
| ES223 | 5475 | M00085100A:H07 | B-30527 |
| ES223 | 5475 | M00085101C:H03 | B-30527 |
| ES223 | 5475 | M00085105D:H02 | B-30527 |
| ES223 | 5475 | M00086323B:C04 | B-30527 |
| ES223 | 5475 | M00084808D:A07 | B-30527 |
| ES223 | 5475 | M00086003D:B10 | B-30527 |
| ES223 | 5475 | M00084787B:D12 | B-30527 |
| ES223 | 5475 | M00085068C:B03 | B-30527 |
| ES223 | 5475 | M00086291A:E10 | B-30527 |
| ES223 | 5475 | M00084807D:F07 | B-30527 |
| ES223 | 5475 | M00084799C:E08 | B-30527 |
| ES223 | 5475 | M00084781A:E05 | B-30527 |
| ES223 | 5475 | M00086324D:D01 | B-30527 |
| ES224 | 5476 | M00085641C:G01 | B-30528 |
| ES224 | 5476 | M00085623B:G11 | B-30528 |
| ES224 | 5476 | M00085805A:G07 | B-30528 |
| ES224 | 5476 | M00085846A:H10 | B-30528 |
| ES224 | 5476 | M00085369D:G04 | B-30528 |
| ES224 | 5476 | M00085817C:G05 | B-30528 |
| ES224 | 5476 | M00085635C:H07 | B-30528 |
| ES224 | 5476 | M00085907B:B11 | B-30528 |
| ES224 | 5476 | M00085650D:E12 | B-30528 |
| ES224 | 5476 | M00085854A:D09 | B-30528 |
| ES224 | 5476 | M00085628B:D03 | B-30528 |
| ES224 | 5476 | M00085393A:D06 | B-30528 |
| ES224 | 5476 | M00085620C:F05 | B-30528 |
| ES224 | 5476 | M00085750B:C10 | B-30528 |
| ES224 | 5476 | M00085773A:F05 | B-30528 |
| ES224 | 5476 | M00085933B:B06 | B-30528 |
| ES224 | 5476 | M00084804B:E01 | B-30528 |
| ES224 | 5476 | M00085337B:F08 | B-30528 |
| ES224 | 5476 | M00085349A:C08 | B-30528 |
| ES224 | 5476 | M00084783C:A09 | B-30528 |
| ES224 | 5476 | M00084782D:D10 | B-30528 |
| ES224 | 5476 | M00085273A:A09 | B-30528 |
| ES224 | 5476 | M00084794B:C01 | B-30528 |
| ES224 | 5476 | M00084780D:D07 | B-30528 |
| ES224 | 5476 | M00085345B:C09 | B-30528 |
| ES224 | 5476 | M00085344D:F01 | B-30528 |
| ES224 | 5476 | M00085747A:B11 | B-30528 |
| ES224 | 5476 | M00085814A:C02 | B-30528 |
| ES224 | 5476 | M00085503D:D05 | B-30528 |
| ES224 | 5476 | M00085304B:D11 | B-30528 |
| ES224 | 5476 | M00085900B:E02 | B-30528 |
| ES224 | 5476 | M00085859B:A11 | B-30528 |
| ES224 | 5476 | M00085860D:H02 | B-30528 |
| ES224 | 5476 | M00085649B:A03 | B-30528 |
| ES224 | 5476 | M00085815C:E11 | B-30528 |

Table 15

| Library ID | CMCC Number | CloneId | NRRL Number |
|---|---|---|---|
| ES224 | 5476 | M00085941B:A06 | B-30528 |
| ES224 | 5476 | M00084800B:H09 | B-30528 |
| ES224 | 5476 | M00085919B:F02 | B-30528 |
| ES224 | 5476 | M00085449A:E02 | B-30528 |
| ES224 | 5476 | M00085344A:G08 | B-30528 |
| ES224 | 5476 | M00085520D:B11 | B-30528 |
| ES224 | 5476 | M00086035A:C11 | B-30528 |
| ES224 | 5476 | M00085955D:H10 | B-30528 |
| ES224 | 5476 | M00086175C:D06 | B-30528 |
| ES224 | 5476 | M00085510B:G12 | B-30528 |
| ES224 | 5476 | M00085927C:G11 | B-30528 |
| ES224 | 5476 | M00085919A:A05 | B-30528 |
| ES224 | 5476 | M00085985C:D02 | B-30528 |
| ES224 | 5476 | M00085934B:E12 | B-30528 |
| ES224 | 5476 | M00085389C:H04 | B-30528 |
| ES224 | 5476 | M00085406A:F03 | B-30528 |
| ES224 | 5476 | M00085826D:B03 | B-30528 |
| ES224 | 5476 | M00085819C:F06 | B-30528 |
| ES224 | 5476 | M00085266B:C06 | B-30528 |
| ES224 | 5476 | M00086112C:A01 | B-30528 |
| ES224 | 5476 | M00086005A:B02 | B-30528 |
| ES224 | 5476 | M00085548D:F01 | B-30528 |
| ES224 | 5476 | M00085809A:E04 | B-30528 |
| ES224 | 5476 | M00085389B:B05 | B-30528 |
| ES224 | 5476 | M00085761C:E07 | B-30528 |
| ES224 | 5476 | M00085454A:G06 | B-30528 |
| ES224 | 5476 | M00085980B:F06 | B-30528 |
| ES224 | 5476 | M00085367B:C02 | B-30528 |
| ES224 | 5476 | M00085922A:E10 | B-30528 |
| ES224 | 5476 | M00085830B:E09 | B-30528 |
| ES224 | 5476 | M00085611C:D09 | B-30528 |
| ES224 | 5476 | M00085810A:A10 | B-30528 |
| ES224 | 5476 | M00085534D:H09 | B-30528 |
| ES224 | 5476 | M00085390D:A03 | B-30528 |
| ES224 | 5476 | M00085419C:H05 | B-30528 |
| ES224 | 5476 | M00085441D:H10 | B-30528 |
| ES224 | 5476 | M00085434C:G06 | B-30528 |
| ES224 | 5476 | M00085428B:G02 | B-30528 |
| ES224 | 5476 | M00086225B:E01 | B-30528 |
| ES224 | 5476 | M00085835B:E11 | B-30528 |
| ES224 | 5476 | M00085590A:G06 | B-30528 |
| ES224 | 5476 | M00086322D:D05 | B-30528 |
| ES224 | 5476 | M00085255D:E12 | B-30528 |
| ES224 | 5476 | M00086259A:F11 | B-30528 |
| ES224 | 5476 | M00086233C:F01 | B-30528 |
| ES224 | 5476 | M00086038A:D03 | B-30528 |
| ES224 | 5476 | M00085569B:C09 | B-30528 |
| ES224 | 5476 | M00084804C:H10 | B-30528 |
| ES224 | 5476 | M00086000A:C05 | B-30528 |
| ES224 | 5476 | M00085605A:D08 | B-30528 |
| ES224 | 5476 | M00083706A:D02 | B-30528 |
| ES224 | 5476 | M00086202C:A07 | B-30528 |
| ES224 | 5476 | M00083691C:E12 | B-30528 |

**Table 15**

| Library ID | CMCC Number | CloneId | NRRL Number |
|---|---|---|---|
| ES224 | 5476 | M00083740C:G01 | B-30528 |
| ES224 | 5476 | M00086159A:F03 | B-30528 |
| ES224 | 5476 | M00085323C:H07 | B-30528 |
| ES224 | 5476 | M00085326B:D08 | B-30528 |
| ES224 | 5476 | M00086270C:D08 | B-30528 |
| ES224 | 5476 | M00086027A:G04 | B-30528 |
| ES224 | 5476 | M00085691A:E06 | B-30528 |
| ES224 | 5476 | M00086276D:F10 | B-30528 |
| ES224 | 5476 | M00086060C:F04 | B-30528 |
| ES224 | 5476 | M00086206A:E10 | B-30528 |
| ES224 | 5476 | M00086087C:D04 | B-30528 |
| ES224 | 5476 | M00083699B:C12 | B-30528 |
| ES224 | 5476 | M00086076B:D02 | B-30528 |
| ES224 | 5476 | M00086288A:B05 | B-30528 |
| ES224 | 5476 | M00085252B:H07 | B-30528 |
| ES224 | 5476 | M00086166D:H04 | B-30528 |
| ES224 | 5476 | M00086184C:D04 | B-30528 |
| ES224 | 5476 | M00085555A:A06 | B-30528 |
| ES224 | 5476 | M00085733D:F08 | B-30528 |
| ES224 | 5476 | M00085887C:B03 | B-30528 |
| ES224 | 5476 | M00086155D:E12 | B-30528 |
| ES224 | 5476 | M00086279C:A08 | B-30528 |
| ES224 | 5476 | M00083710D:B09 | B-30528 |
| ES224 | 5476 | M00086228A:F11 | B-30528 |
| ES224 | 5476 | M00086145A:F07 | B-30528 |
| ES224 | 5476 | M00085100B:C12 | B-30528 |
| ES224 | 5476 | M00085287C:G08 | B-30528 |
| ES224 | 5476 | M00083729C:F10 | B-30528 |
| ES224 | 5476 | M00083692C:D09 | B-30528 |
| ES224 | 5476 | M00086090B:B09 | B-30528 |
| ES224 | 5476 | M00086120A:D11 | B-30528 |
| ES224 | 5476 | M00086031C:G11 | B-30528 |
| ES224 | 5476 | M00085294D:G03 | B-30528 |
| ES224 | 5476 | M00086114D:G11 | B-30528 |
| ES224 | 5476 | M00085720D:A03 | B-30528 |
| ES224 | 5476 | M00085262A:A02 | B-30528 |
| ES224 | 5476 | M00085732A:G04 | B-30528 |
| ES224 | 5476 | M00086159B:E04 | B-30528 |
| ES224 | 5476 | M00084781A:H09 | B-30528 |
| ES224 | 5476 | M00086235A:F05 | B-30528 |
| ES224 | 5476 | M00086097C:B10 | B-30528 |
| ES224 | 5476 | M00084779B:D03 | B-30528 |
| ES224 | 5476 | M00086085D:F09 | B-30528 |
| ES224 | 5476 | M00086286C:H02 | B-30528 |
| ES224 | 5476 | M00083733B:D08 | B-30528 |
| ES224 | 5476 | M00085322D:G05 | B-30528 |
| ES224 | 5476 | M00085071B:D07 | B-30528 |
| ES224 | 5476 | M00085707A:A04 | B-30528 |
| ES224 | 5476 | M00083736A:D10 | B-30528 |
| ES224 | 5476 | M00085301C:H11 | B-30528 |
| ES224 | 5476 | M00085066D:A05 | B-30528 |
| ES224 | 5476 | M00086294D:F08 | B-30528 |
| ES224 | 5476 | M00084773C:D04 | B-30528 |

**Table 15**

| Library ID | CMCC Number | CloneId | NRRL Number |
|---|---|---|---|
| ES224 | 5476 | M00085280D:F06 | B-30528 |
| ES224 | 5476 | M00086302A:E06 | B-30528 |
| ES224 | 5476 | M00085084A:E12 | B-30528 |
| ES224 | 5476 | M00085105C:D01 | B-30528 |
| ES224 | 5476 | M00085297A:A11 | B-30528 |
| ES224 | 5476 | M00085076C:H01 | B-30528 |
| ES224 | 5476 | M00086286B:F08 | B-30528 |
| ES224 | 5476 | M00085107D:H08 | B-30528 |
| ES224 | 5476 | M00085092D:D09 | B-30528 |
| ES225 | 5477 | M00086103D:E08 | B-30529 |
| ES225 | 5477 | M00086272D:E04 | B-30529 |
| ES225 | 5477 | M00085449B:G12 | B-30529 |
| ES225 | 5477 | M00085832D:G02 | B-30529 |
| ES225 | 5477 | M00085827C:F03 | B-30529 |
| ES225 | 5477 | M00086328B:G12 | B-30529 |
| ES225 | 5477 | M00085820D:D02 | B-30529 |
| ES225 | 5477 | M00085522C:E05 | B-30529 |
| ES225 | 5477 | M00086178A:D07 | B-30529 |
| ES225 | 5477 | M00085651A:A01 | B-30529 |
| ES225 | 5477 | M00085814B:G08 | B-30529 |
| ES225 | 5477 | M00086149A:F06 | B-30529 |
| ES225 | 5477 | M00085754C:A12 | B-30529 |
| ES225 | 5477 | M00085383C:C05 | B-30529 |
| ES225 | 5477 | M00086206C:C01 | B-30529 |
| ES225 | 5477 | M00085982D:C06 | B-30529 |
| ES225 | 5477 | M00085590B:F11 | B-30529 |
| ES225 | 5477 | M00086161B:H08 | B-30529 |
| ES225 | 5477 | M00086277B:E06 | B-30529 |
| ES225 | 5477 | M00085305B:F10 | B-30529 |
| ES225 | 5477 | M00086081B:A06 | B-30529 |
| ES225 | 5477 | M00086053B:F01 | B-30529 |
| ES225 | 5477 | M00085259A:H05 | B-30529 |
| ES225 | 5477 | M00083694C:F09 | B-30529 |
| ES225 | 5477 | M00085643D:F06 | B-30529 |
| ES225 | 5477 | M00086209B:H12 | B-30529 |
| ES225 | 5477 | M00085860B:D10 | B-30529 |
| ES225 | 5477 | M00086178D:H12 | B-30529 |
| ES225 | 5477 | M00085557B:B02 | B-30529 |
| ES225 | 5477 | M00086184C:C10 | B-30529 |
| ES225 | 5477 | M00086227B:E06 | B-30529 |
| ES225 | 5477 | M00085603A:E01 | B-30529 |
| ES225 | 5477 | M00086233D:A03 | B-30529 |
| ES225 | 5477 | M00085817D:C04 | B-30529 |
| ES225 | 5477 | M00086157D:D03 | B-30529 |
| ES225 | 5477 | M00085583A:E06 | B-30529 |
| ES225 | 5477 | M00086136C:B06 | B-30529 |
| ES225 | 5477 | M00085626B:B11 | B-30529 |
| ES225 | 5477 | M00086048D:H08 | B-30529 |
| ES225 | 5477 | M00086010B:B05 | B-30529 |
| ES225 | 5477 | M00083735C:H12 | B-30529 |
| ES225 | 5477 | M00083722B:A07 | B-30529 |
| ES225 | 5477 | M00085745C:C03 | B-30529 |
| ES225 | 5477 | M00085809A:C05 | B-30529 |

Table 15

| Library ID | CMCC Number | CloneId | NRRL Number |
|---|---|---|---|
| ES225 | 5477 | M00085694D:D06 | B-30529 |
| ES225 | 5477 | M00085786D:G12 | B-30529 |
| ES225 | 5477 | M00085764B:H12 | B-30529 |
| ES225 | 5477 | M00086000C:B08 | B-30529 |
| ES225 | 5477 | M00085817B:B08 | B-30529 |
| ES225 | 5477 | M00085806B:A10 | B-30529 |
| ES225 | 5477 | M00086081D:D09 | B-30529 |
| ES225 | 5477 | M00085808C:E12 | B-30529 |
| ES225 | 5477 | M00085336C:G09 | B-30529 |
| ES225 | 5477 | M00085620B:D08 | B-30529 |
| ES225 | 5477 | M00085721A:G03 | B-30529 |
| ES225 | 5477 | M00086128D:H10 | B-30529 |
| ES225 | 5477 | M00085361A:A09 | B-30529 |
| ES225 | 5477 | M00085714C:G03 | B-30529 |
| ES225 | 5477 | M00085741C:D06 | B-30529 |
| ES225 | 5477 | M00085722A:A06 | B-30529 |
| ES225 | 5477 | M00083704C:C04 | B-30529 |
| ES225 | 5477 | M00085549D:G03 | B-30529 |
| ES225 | 5477 | M00086143B:B08 | B-30529 |
| ES225 | 5477 | M00085926C:C06 | B-30529 |
| ES225 | 5477 | M00085980A:G10 | B-30529 |
| ES225 | 5477 | M00085625B:F01 | B-30529 |
| ES225 | 5477 | M00086128A:D09 | B-30529 |
| ES225 | 5477 | M00085393D:F12 | B-30529 |
| ES225 | 5477 | M00085935D:H04 | B-30529 |
| ES225 | 5477 | M00086159D:D01 | B-30529 |
| ES225 | 5477 | M00085597C:C03 | B-30529 |
| ES225 | 5477 | M00085259D:B06 | B-30529 |
| ES225 | 5477 | M00086015D:G04 | B-30529 |
| ES225 | 5477 | M00085255D:B09 | B-30529 |
| ES225 | 5477 | M00083715A:B11 | B-30529 |
| ES225 | 5477 | M00085959B:D04 | B-30529 |
| ES225 | 5477 | M00085380B:E10 | B-30529 |
| ES225 | 5477 | M00085100A:A12 | B-30529 |
| ES225 | 5477 | M00085350D:G05 | B-30529 |
| ES225 | 5477 | M00086301A:D04 | B-30529 |
| ES225 | 5477 | M00085891D:E07 | B-30529 |
| ES225 | 5477 | M00085108A:C12 | B-30529 |
| ES225 | 5477 | M00085085C:D10 | B-30529 |
| ES225 | 5477 | M00085104C:A10 | B-30529 |
| ES225 | 5477 | M00084805D:E02 | B-30529 |
| ES225 | 5477 | M00084789D:B10 | B-30529 |
| ES225 | 5477 | M00085277D:C11 | B-30529 |
| ES225 | 5477 | M00085647A:E04 | B-30529 |
| ES225 | 5477 | M00085886C:F05 | B-30529 |
| ES225 | 5477 | M00085444C:C02 | B-30529 |
| ES225 | 5477 | M00085415A:D12 | B-30529 |
| ES225 | 5477 | M00085435A:B11 | B-30529 |
| ES225 | 5477 | M00085282C:F05 | B-30529 |
| ES225 | 5477 | M00084784B:A02 | B-30529 |
| ES225 | 5477 | M00085944A:F04 | B-30529 |
| ES225 | 5477 | M00085804B:F09 | B-30529 |
| ES225 | 5477 | M00085339C:C04 | B-30529 |

**Table 15**

| Library ID | CMCC Number | CloneId | NRRL Number |
|---|---|---|---|
| ES225 | 5477 | M00085968B:F09 | B-30529 |
| ES225 | 5477 | M00084779D:H10 | B-30529 |
| ES225 | 5477 | M00085381C:A05 | B-30529 |
| ES225 | 5477 | M00084783C:G08 | B-30529 |
| ES225 | 5477 | M00085333B:B10 | B-30529 |
| ES225 | 5477 | M00085068B:A07 | B-30529 |
| ES225 | 5477 | M00084773C:H08 | B-30529 |
| ES225 | 5477 | M00086020C:E09 | B-30529 |
| ES225 | 5477 | M00085273B:F08 | B-30529 |
| ES225 | 5477 | M00085361D:F06 | B-30529 |
| ES225 | 5477 | M00084780C:F08 | B-30529 |
| ES225 | 5477 | M00085302C:B06 | B-30529 |
| ES225 | 5477 | M00085988D:F09 | B-30529 |
| ES225 | 5477 | M00083738C:D05 | B-30529 |
| ES225 | 5477 | M00085331C:C07 | B-30529 |
| ES225 | 5477 | M00085541C:F06 | B-30529 |
| ES225 | 5477 | M00086327B:D10 | B-30529 |
| ES225 | 5477 | M00085917C:A04 | B-30529 |
| ES225 | 5477 | M00086310A:F04 | B-30529 |
| ES225 | 5477 | M00085510C:A07 | B-30529 |
| ES225 | 5477 | M00085296D:H10 | B-30529 |
| ES225 | 5477 | M00086085A:G05 | B-30529 |
| ES225 | 5477 | M00085299B:G01 | B-30529 |
| ES225 | 5477 | M00085503D:G05 | B-30529 |
| ES225 | 5477 | M00085260C:A10 | B-30529 |
| ES225 | 5477 | M00086282A:B10 | B-30529 |
| ES225 | 5477 | M00085837C:H08 | B-30529 |
| ES225 | 5477 | M00085630B:B09 | B-30529 |
| ES225 | 5477 | M00086279D:C07 | B-30529 |
| ES225 | 5477 | M00085263C:F12 | B-30529 |
| ES225 | 5477 | M00086294B:E11 | B-30529 |
| ES225 | 5477 | M00086322A:E02 | B-30529 |
| ES225 | 5477 | M00085854C:E06 | B-30529 |
| ES225 | 5477 | M00086272D:H11 | B-30529 |
| ES225 | 5477 | M00085422D:D07 | B-30529 |
| ES225 | 5477 | M00085844C:H11 | B-30529 |
| ES225 | 5477 | M00085288C:C09 | B-30529 |
| ES225 | 5477 | M00085082C:B04 | B-30529 |
| ES225 | 5477 | M00085103D:H12 | B-30529 |
| ES225 | 5477 | M00086336B:A08 | B-30529 |
| ES225 | 5477 | M00084773C:F08 | B-30529 |
| ES225 | 5477 | M00085896D:A09 | B-30529 |
| ES225 | 5477 | M00085324B:F10 | B-30529 |
| ES225 | 5477 | M00085267B:D06 | B-30529 |
| ES225 | 5477 | M00085430C:E04 | B-30529 |
| ES225 | 5477 | M00085312C:B09 | B-30529 |
| ES225 | 5477 | M00085074B:A07 | B-30529 |
| ES225 | 5477 | M00085918D:C11 | B-30529 |
| ES225 | 5477 | M00085341C:H08 | B-30529 |
| ES225 | 5477 | M00084791D:D01 | B-30529 |
| ES225 | 5477 | M00085471B:H09 | B-30529 |
| ES225 | 5477 | M00085893B:D08 | B-30529 |
| ES225 | 5477 | M00084781A:A05 | B-30529 |

**Table 15**

| Library ID | CMCC Number | CloneId | NRRL Number |
|---|---|---|---|
| ES226 | 5478 | M00084956B:B05 | B-30581 |
| ES226 | 5478 | M00084954D:D12 | B-30581 |
| ES226 | 5478 | M00084948B:F04 | B-30581 |
| ES226 | 5478 | M00084950D:F05 | B-30581 |
| ES226 | 5478 | M00084954D:E01 | B-30581 |
| ES226 | 5478 | M00084941D:C10 | B-30581 |
| ES226 | 5478 | M00084950D:A06 | B-30581 |
| ES226 | 5478 | M00084941D:H02 | B-30581 |
| ES226 | 5478 | M00084954C:B12 | B-30581 |
| ES226 | 5478 | M00084955A:E08 | B-30581 |
| ES226 | 5478 | M00084954D:A05 | B-30581 |
| ES226 | 5478 | M00084951A:D04 | B-30581 |
| ES226 | 5478 | M00084954C:A03 | B-30581 |

# Table 16

| SEQ ID | Seq Name | SpotID | Breast Cancer Tumor/Normal >=2x | Breast Cancer Patients | Colon Cancer Tumor/Normal >=2x | Colon Cancer Patients | Prostate Cancer Tumor/Normal >=2x | Prostate Cancer Patients | Colon Unmatched Met/Normal >=2x | Colon Unmatched Met Patients | Colon Matched Met/Normal >=2x | Colon Matched Met Patients | Colon Matched Met/Tumor >=2x | Colon Match Met Patients |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 3541.A16.GZ43_505167 | 58648 | | 23 | 26.09 | | | | | | | | | |
| 18 | 3538.K12.GZ43_504729 | 56775 | | 23 | 26.00 | | | | | | | | | |
| 48 | 3544.A17.GZ43_505567 | 56939 | | | | | | | | | | | 22.22 | 18 |
| 72 | 3544.L13.GZ43_505514 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 129 | 3550.D16.GZ43_506322 | 42108 | | | 44.74 | 76 | | | 63.64 | 33 | 52.78 | 36 | | |
| 191 | 3553.J14.GZ43_506680 | 60233 | | | 47.37 | 19 | 37.11 | 97 | | | 41.18 | 17 | | |
| 198 | 3553.K03.GZ43_506505 | 55773 | 26.09 | 23 | 21.05 | 19 | 22.45 | 98 | | | | | | |
| 241 | 3556.C15.GZ43_507073 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 259 | 3556.J14.GZ43_507064 | 60233 | | | 47.37 | 19 | 37.11 | 97 | | | 41.18 | 17 | | |
| 269 | 3556.M02.GZ43_506875 | 42108 | | | 44.74 | 76 | | | 63.64 | 33 | 52.78 | 36 | | |
| 275 | 3556.N06.GZ43_506940 | 58440 | | | | | 20.41 | 98 | | | | | | |
| 334 | 3562.I02.GZ43_507639 | 58075 | | | | | 21.43 | 98 | | | | | | |
| 466 | 3574.F10.GZ43_509318 | 58075 | | | | | 21.43 | 98 | | | | | | |
| 470 | 3574.G11.GZ43_509335 | 56782 | 21.74 | 23 | | | | | | | | | | |
| 472 | 3574.I02.GZ43_509193 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 621 | 3590.D19.GZ43_512389 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 635 | 3590.J01.GZ43_512107 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 644 | 3590.L10.GZ43_512253 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 659 | 3596.E08.GZ43_512598 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 671 | 3596.K14.GZ43_512700 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 681 | 3596.O10.GZ43_512640 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 685 | 3596.P07.GZ43_512593 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 708 | 3599.K23.GZ43_513228 | 57429 | | | | | 34.69 | 98 | | | | | | |
| 708 | 3599.K23.GZ43_513228 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 713 | 3599.M24.GZ43_513246 | 35065 | | | 24.00 | 75 | | | 21.21 | 33 | | | | |
| 718 | 3599.O06.GZ43_512960 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 721 | 3602.A09.GZ43_513378 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 740 | 3602.K06.GZ43_513340 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 750 | 3605.I19.gz43_513930 | 56753 | | | | | 20.41 | 98 | | | | | | |
| 790 | 3611.I04.gz43_514458 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 798 | 3611.L22.gz43_514749 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 820 | 3614.P11.gz43_514961 | 56775 | 26.09 | 23 | | | | | | | | | | |
| 822 | 3617.B16.gz43_515411 | 1368 | | | 34.29 | 35 | | | 56.67 | 30 | 28.57 | 7 | | |
| | 3617.B16.gz43_515411 | 25873 | | | 50.00 | 76 | | | 57.58 | 33 | 55.56 | 36 | | |
| | 3617.B16.gz43_515411 | 26658 | | | 59.21 | 76 | | | 66.67 | 33 | 44.44 | 36 | | |
| 825 | 3617.H16.gz43_515417 | 59904 | | | | | 25.77 | 97 | | | | | | |
| 826 | 3617.I01.gz43_515178 | 57008 | | | | | 23.47 | 98 | | | | | | |
| 833 | 3617.N14.gz43_515391 | 57651 | | | | | 21.43 | 98 | | | | | | |
| 835 | 3617.P11.gz43_515345 | 56753 | | | | | 20.41 | 98 | | | | | | |
| 836 | 3617.P12.gz43_515361 | 60100 | | | | | 31.96 | 97 | | | | | | |

# Table 16

| SEQ ID | Seq Name | SpotID | Breast Cancer Tumor/Normal >=2x | Breast Cancer Patients | Colon Cancer Tumor/Normal >=2x | Colon Cancer Patients | Prostate Cancer Tumor/Normal >=2x | Prostate Cancer Patients | Colon Unmatched Met/Normal >=2x | Colon Unmatched Met Patients | Colon Matched Met/Normal >=2x | Colon Matched Met Patients | Colon Matched Met/Tumor >=2x | Colon Match Met Patients |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 838 | 3620.B03.gz43_515810 | 57651 | | | | | 21.43 | 98 | | | | | | |
| 846 | 3620.G17.gz43_516039 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 866 | 3623.N23.gz43_516526 | 27078 | | | 28.95 | 76 | | | 15.15 | 33 | | | | |
| 872 | 3626.G01.gz43_516551 | 33958 | 39.13 | 23 | | | 24.51 | 102 | 18.18 | 33 | | | | |
| | 3626.G01.gz43_516551 | 35113 | 39.13 | 23 | | | 23.53 | 102 | 15.15 | 33 | | | | |
| | 3626.G01.gz43_516551 | 58921 | 30.43 | 23 | | | 22.45 | 98 | | | | | | |
| 876 | 3626.M15.gz43_516781 | 59829 | 26.09 | 23 | 36.84 | 19 | | | | | 47.06 | 17 | | |
| 897 | 3632.G01.gz43_517319 | 25933 | | | | | 39.22 | 102 | 0.00 | 33 | | | | |
| 899 | 3632.K20.gz43_517627 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 901 | 3632.M13.gz43_517517 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 902 | 3632.M19.gz43_517613 | 57429 | | | | | 34.69 | 98 | | | | | | |
| | 3632.M19.gz43_517613 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 909 | 3635.A13.gz43_517889 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 930 | 3638.L10.gz43_518236 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 941 | 3643.I24.gz43_518841 | 59904 | | | | | 25.77 | 97 | | | | | | |
| 984 | 3661.K22.gz43_519717 | 56753 | | | | | 20.41 | 98 | | | | | | |
| 1043 | 3664.K19.gz43_520821 | 25933 | | | | | 39.22 | 102 | 0.00 | 33 | | | | |
| 1046 | 3664.P12.gz43_520714 | 57008 | | | | | 23.47 | 98 | | | | | | |
| | 3664.P12.gz43_520714 | 57797 | | | | | 21.43 | 98 | | | | | | |
| 1081 | 3666.L23.gz43_521654 | 53114 | | | | | 26.47 | 102 | 0.00 | 33 | | | | |
| 1091 | 3754.B08.gz43_532950 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 1124 | 3756.A02.gz43_533237 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 1133 | 3756.C16.gz43_533463 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 1139 | 3756.E12.gz43_533401 | 57651 | | | | | 21.43 | 98 | | | | | | |
| 1145 | 3756.G14.gz43_533435 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 1173 | 3759.K05.gz43_533679 | 59904 | | | | | 25.77 | 97 | | | | | | |
| 1192 | 3762.A20.gz43_534293 | 60233 | | | 47.37 | 19 | 37.11 | 97 | | | 41.18 | 17 | | |
| 1215 | 3762.L18.gz43_534272 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 1221 | Clu8293.con_1 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 1239 | Clu403488.con_1 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 1276 | Clu609914.con_1 | 24511 | | | | | | | 24.24 | 33 | 26.09 | 23 | | |
| 1278 | Clu621702.con_1 | 35065 | | | 24.00 | 75 | | | 21.21 | 33 | | | | |
| 1294 | Clu733840.con_1 | 24511 | | | | | | | 24.24 | 33 | 26.09 | 23 | | |
| 1302 | Clu777670.con_1 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 1308 | Clu854573.con_1 | 57429 | | | | | 34.69 | 98 | | | | | | |
| | Clu854573.con_1 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 1326 | Clu1053799.con_1 | 58075 | | | | | 21.43 | 98 | | | | | | |
| 1332 | Clu1054813.con_1 | 60233 | | | 47.37 | 19 | 37.11 | 97 | | | 41.18 | 17 | | |
| 1338 | Clu1055326.con_1 | 25933 | | | | | 39.22 | 102 | 0.00 | 33 | | | | |
| 1359 | Clu1088930.con_1 | 27078 | | | 28.95 | 76 | | | 15.15 | 33 | | | | |

# Table 16

| SEQ ID | Seq Name | SpotID | Breast Cancer Tumor/Normal >=2x | Breast Cancer Patients | Colon Cancer Tumor/Normal >=2x | Colon Cancer Patients | Prostate Cancer Tumor/Normal >=2x | Prostate Cancer Patients | Colon Unmatched Met/Normal >=2x | Colon Unmatched Met Patients | Colon Matched Met/Normal >=2x | Colon Matched Met Patients | Colon Matched Met/Tumor >=2x | Colon Match Met Patients |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1389 | Clu1224379.con_1 | 42108 | | | 44.74 | 76 | | | 63.64 | 33 | 52.78 | 36 | | |
| 1408 | Clu1228277.con_1 | 60100 | | | | | 31.96 | 97 | | | | | | |
| 1413 | Clu1259069.con_2 | 25844 | | | | | 51.96 | 102 | 0.00 | 33 | | | | |
| | Clu1259069.con_2 | 28996 | | | | | 49.02 | 102 | 0.00 | 33 | | | | |
| 1416 | Clu1292262.con_1 | 56753 | | | | | 20.41 | 98 | | | | | | |
| 1421 | Clu1292436.con_1 | 59904 | | | | | 25.77 | 97 | | | | | | |
| 1440 | NTN_007592S2.3_10 | 53100 | | | 25.33 | 75 | | | 6.06 | 33 | 28.57 | 35 | | |
| 1451 | NTN_009296S1.3_1 | 1368 | | | 34.29 | 35 | | | 56.67 | 30 | 28.57 | 7 | | |
| | NTN_009296S1.3_1 | 25873 | | | 50.00 | 76 | | | 57.58 | 33 | 55.56 | 36 | | |
| 1451 | NTN_009296S1.3_1 | 26658 | | | 59.21 | 76 | | | 66.67 | 33 | 44.44 | 36 | | |
| 1452 | NTN_009296S3.3_2 | 56753 | | | | | 20.41 | 98 | | | | | | |
| 1467 | NTN_011512S51.3_3 | 35086 | | | 22.67 | 75 | | | 9.09 | 33 | 25.00 | 36 | | |
| | NTN_011512S51.3_3 | 36824 | | | 21.33 | 75 | | | 12.12 | 33 | 33.33 | 36 | | |
| 1468 | NTN_017582S2.3_6 | 26345 | | | 69.74 | 76 | | | 78.79 | 33 | 66.67 | 36 | | |
| 1482 | NTN_025842S13.2_1 | 27078 | | | 28.95 | 76 | | | 15.15 | 33 | | | | |

We Claim:

1. An isolated polynucleotide comprising a nucleotide sequence which hybridizes under stringent conditions to a sequence selected from the group consisting of SEQ ID NOS: 1-1485, or complement thereof

2. An isolated polynucleotide comprising at least 15 contiguous nucleotides of a nucleotide sequence having at least 90% sequence identity to a sequence selected from the group consisting of SEQ ID NOS:1-1485, or complement thereof.

3. An isolated polynucleotide comprising at least 15 contiguous nucleotides of a nucleotide sequence selected from the group consisting of SEQ ID NOS:1-1485, or complement thereof.

4. The isolated polynucleotide of any one of claims 1-3, wherein the polynucleotide comprises at least 100 contiguous nucleotides of the nucleotide sequence or complement thereof.

5. The isolated polynucleotide of any one of claims 1-4, wherein the polynucleotide comprises at least 200 contiguous nucleotides of the selected nucleotide sequence or complement thereof.

6. An isolated polynucleotide comprising a nucleotide sequence of at least 90% sequence identity to a sequence selected from the group consisting of: SEQ ID NOS:1-1485 or complement therefore.

7. The isolated polynucleotide of claim 6, wherein the polynucleotide comprises a nucleotide sequence of at least 95% sequence identity to the selected nucleotide sequence.

8. The isolated polynucleotide of claim 6, wherein the polynucleotide comprises a nucleotide sequence that is identical to the selected nucleotide sequence.

9. A polynucleotide comprising a nucleotide sequence of an insert contained in a clone deposited as NRRL Accession No. B-30523, B-30524, B-30525, B-30526, B-30527, B-30528, B-30529, or B-30581.

10. An isolated cDNA obtained by the process of amplification using a polynucleotide comprising at least 15 contiguous nucleotides of a nucleotide sequence selected from the group consisting of SEQ ID NOS:1-1485.

11. The isolated cDNA of claim 10, wherein the polynucleotide comprises at least 25 contiguous nucleotides of the selected nucleotide sequence.

12. The isolated cDNA of claim 10, wherein the polynucleotide comprises at least 100 contiguous nucleotides of the selected nucleotide sequence.

13. The isolated cDNA of claims 10, 11, or 12, wherein amplification is by polymerase chain reaction (PCR) amplification.

14. An isolated recombinant host cell containing the polynucleotide according to claims 1, 2, 3, 6, 9, or 10.

15. An isolated vector comprising the polynucleotide according to claims 1, 2, 3, 6, 9, or 10.

16. A method for producing a polypeptide, the method comprising the steps of:

culturing a recombinant host cell containing the polynucleotide according to claims 1, 2, 3, 6, 9, or 10, said culturing being under conditions suitable for the expression of an encoded polypeptide; and

recovering the polypeptide from the host cell culture.

17. An isolated polypeptide encoded by the polynucleotide according to claims 1, 2, 3, 6, 9, or 10.

18. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NOS:1486-1542.

19. An antibody that specifically binds the polypeptide of claim 17 or 18.

20. A library of polynucleotides, wherein at least one of the polynucleotides comprises the sequence information of the polynucleotide according to claims 1, 2, 3, 6, 9, or 10.

21. The library of claim 20, wherein the library is provided on a nucleic acid array.

22. The library of claim 20, wherein the library is provided in a computer-readable format.

23. A method for detecting a cancerous cell, said method comprising:

detecting a level of a product of a gene in a test sample obtained from a cell of a subject, wherein said gene is identified by a sequence having at least 80% sequence identity to a sequence selected from a group consisting of SEQ ID NOS:1-1485, or a fragment thereof; and,

comparing the level of said product to a control level of said gene product, wherein the presence of a cancerous cell is indicated by detection of said level and comparison to a control level of said gene product.

24. The method of claim 23, wherein said gene product is nucleic acid.

25. The method of claim 23, wherein said detecting step uses a polymerase chain reaction.

26. The method of claim 23, wherein said detecting step uses hybridization.

27. The method of claim 23, wherein said sample is a sample of prostate, colon or breast tissue.

28. A method for inhibiting a cancerous phenotype of a cell, said method comprising:

contacting a mammalian cell with an agent for inhibition of a product of a gene, wherein said gene is identified by a sequence having at least 80% sequence identity to a sequence selected from a group consisting of SEQ ID NOS:1-1485, or a fragment thereof.

29. The method of claim 28, wherein said cancerous phenotype is aberrant cellular proliferation relative to a normal cell.

30. A method of treating a subject with cancer, said method comprising:

administering to a subject a pharmaceutically effective amount of an agent,

wherein said agent modulates the activity of a product of a gene identified by a sequence having at least 80% sequence identity to a sequence selected from a group consisting of SEQ ID NOS:1-1485, or a fragment thereof.

31. A method for identifying an agent that modulates a biological activity of a gene product differentially expressed in a cancerous cell as compared to a normal cell, said method comprising:

contacting a candidate agent with a product of a gene encoded by a gene defined by a sequence having at least 80% sequence identity to a sequence selected from a group consisting of SEQ ID NOS:1-1485, or a fragment thereof; and

detecting modulation of a biological activity of the gene product relative to a level of biological activity of the gene product in the absence of the candidate agent.